J. Fluid Mech. (2016), *vol.* 788, *pp.* 492–520.
doi:10.1017/jfm.2015.706

Resonant fast–slow interactions and breakdown of quasi-geostrophy in rotating shallow water

Jim Thomas†

Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA

(Received 16 March 2015; revised 10 October 2015; accepted 26 November 2015; first published online 8 January 2016)

In this paper we investigate the possibility of fast waves affecting the evolution of slow balanced dynamics in the regime $Ro \sim Fr \ll 1$ of a rotating shallow water system, where Ro and Fr are the Rossby and Froude numbers respectively. The problem is set up as an initial value problem with unbalanced initial data. The method of multiple time scale asymptotic analysis is used to derive an evolution equation for the slow dynamics that holds for $t \lesssim 1/(fRo^2)$, f being the inertial frequency. This slow evolution equation is affected by the fast waves and thus does not form a closed system. Furthermore, it is shown that energy and enstrophy exchange can take place between the slow and fast dynamics. As a consequence, the quasi-geostrophic ideology of describing the slow dynamics of the balanced flow without any information on the fast modes breaks down. Further analysis is carried out in a doubly periodic domain for a few geostrophic and wave modes. A simple set of slowly evolving amplitude equations is then derived using resonant wave interaction theory to demonstrate that significant wave-balanced flow interactions can take place in the long-time limit. In this reduced system consisting of two geostrophic modes and two wave modes, the presence of waves considerably affects the interactions between the geostrophic modes, the waves acting as a catalyst in promoting energetic interactions among geostrophic modes.

Key words: quasi-geostrophic flows, wave–turbulence interactions, waves in rotating fluids

1. Introduction

The large-scale flow in the atmosphere and the ocean appears to be approximately in geostrophic balance and thus divergence free (Gill 1982). The scale separation between fast gravity waves and slowly evolving balanced flow supports this observation. In a celebrated article, Rossby (1938) used the one-dimensional rotating shallow water equations (RSW hereafter) to illustrate the phenomenon of geostrophic adjustment, where an unbalanced initial state attains balance after a long time due to the radiation of waves away from the local region of excitation. This remarkable result was due to the fact that the linear potential vorticity (hereafter PV) is independent of the wave activity and does not evolve in inertial time. Thus, the state of the system after a long time can be obtained by equating the initial and final PV, and then inverting the PV to get the velocity and pressure (or height) field. This was followed

† Email address for correspondence: jthomas@cims.nyu.edu

by a series of works that addressed the problem in different contexts and greater detail (see Blumen (1972) and references therein). The scale separation between fast waves and balanced flow was used by Charney (1948) and Obukhov (1949) independently to derive the quasi-geostrophic (QG) equation. The most remarkable feature of QG theory is that one can track the slow evolution of the balanced flow without having any information on the dynamics of fast waves in the system.

This was followed by the introduction of the idea of a slow manifold (Leith 1980; Lorenz 1980), whereby the evolution of the balanced flow is obtained by projecting the dynamics onto a subset of the whole space. The slow manifold – its existence and exact definition – has been debated since its introduction (Lorenz 1986; Vautard & Legras 1986; Lorenz & Krishnamurty 1987). A major challenge to its existence is the phenomenon of spontaneous generation, by which an initially balanced flow can develop wave activity in finite time. Although investigations in this direction seem to indicate the non-existence of the slow manifold in its strict sense as a manifold of zero thickness in state space, at present, conventional wisdom seems to have established that in the regime $Ro \sim Fr \ll 1$, where Ro (ratio of rotational time scales to advective time scales) and Fr (ratio of flow velocity to wave speed) refer to the Rossby and Froude numbers respectively, often called the QG regime, the amplitude of the waves excited spontaneously is exponentially small (see Vanneste (2013) and references therein and Vanneste & Yavneh (2004) for a specific example).

The problem we address in this work is not the evolution of balanced initial data and subsequent spontaneous generation of waves, but rather the evolution of unbalanced initial data in RSW and the resulting interaction between the waves and the balanced flow, both being of comparable magnitude. In such a setting, for $\epsilon = Ro \sim Fr \ll 1$, several works have used formal asymptotic methods to derive the QG equation in RSW more systematically and establish the absence of the fast motion affecting the balanced flow up to $O(\epsilon)$. Majda & Embid (1998) in spectral space and Dewar & Killworth (1995) in physical space illustrated that the geostrophic field is unaffected by the waves for time scales $t \lesssim 1/(\epsilon f)$, f being the inertial frequency. Further numerical simulations by Dewar & Killworth showed a lack of energy exchange between the waves and the geostrophic modes. From a turbulence perspective, Warn (1986) used statistical mechanics investigations on RSW to conclude that after a long time most of the energy ends up with the waves. Warn speculated about the possibility of energy exchange between waves and geostrophic modes in the long-time limit. In a low-order truncated forced-dissipative setting, the presence of gravity waves even after long times was pointed out by Warn & Menard (1986). Careful numerical experiments on RSW by Farge & Sadourny (1989), however, did not find any energy exchange between the waves and the balanced flow, although they found that the presence of inertio-gravitational energy in the system prevented the inverse energy cascade of the geostrophic flow. Farge & Sadourny did point out the possibility that artificial dissipation, used for numerical stability in their simulations, could have prevented the system from relaxing to statistical equilibrium. The inertio-gravity waves in RSW, due to their specific form of dispersion relation, do not allow triad interactions (Babin, Mahalov & Nicolaenko 1997; Majda 2002). Hence, there are no resonant wave triads although the geostrophic modes can catalyse interaction between wave modes. This was investigated in detail by Ward & Dewar (2010), who showed that a coherent wave energy distribution can be scattered and inhomogenized by the geostrophic modes. Numerical simulations of reduced models and a full RSW system did not find any energy exchange between the waves and the balanced flow. In an unbounded domain with compactly supported unbalanced initial data, which is the classical setting for the geostrophic adjustment problem, the waves do not influence the balanced flow even with nonlinear interactions due to their rapid propagation away from the initial region. Reznik, Zeitlin & Ben Jelloul (2001) (RZB hereafter) used multiple time scale asymptotics in such a setting to argue that the waves do not affect the balanced flow up to $t \sim 1/(\epsilon^2 f)$ in RSW. Reznik & Grimshaw (2002) further investigated the adjustment problem in the half-plane, i.e. in the presence of a boundary at $y=0$, a setting that can support Kelvin waves. For different sets of initial conditions considered, except in the case of periodic initial data in x where the Kelvin waves were seen to affect the evolution of balanced dynamics, geostrophic adjustment leading to the slow evolution of the balanced flow unaffected by the waves was observed in all other cases. In a barotropic rotating fluid, without making the classical traditional approximation, i.e. neglecting the projection of Coriolis force on the rotating plane and in the absence of hydrostatic balance, Reznik (2014) showed that gyroscopic waves can affect the evolution of balanced dynamics in the long-time limit.

At present, conventional wisdom seems to be that in the regime $Ro \sim Fr \ll 1$ of RSW with unbalanced initial data that can excite inertio-gravity waves and balanced flow of equal magnitude, the geostrophically balanced flow evolves slowly, as described by the QG equation, unaffected by the waves, while the balanced flow can catalyse interactions between the fast waves. The possibility of fast waves influencing the balanced flow in this regime still remains an open question, in spite of the many advances that have been made in geophysical flows. The differences in the various types of models used (forced-dissipative, truncated-inviscid, etc.) in the numerical works previously described also lead to different answers. The only theoretical work that has investigated the state of unbalanced initial conditions for very long time scales ($t \sim 1/(\epsilon^2 f)$) in this particular parameter regime of RSW with an eye on the possibility of inertio-gravity waves affecting balanced dynamics, to the best of the present author's knowledge, is RZB, their set-up being an unbounded domain with compact initial data, leading to geostrophic adjustment. This sets the main motivation for the present theoretical work aimed at investigating the possibility of fast waves affecting the balanced flow by investigating the evolution of unbalanced initial data in RSW.

We use the method of multiscale asymptotics to investigate fast–slow interactions in RSW. The initial conditions are arbitrary or unbalanced with data that would project on fast modes and balanced flow, the fast part consisting of inertio-gravity waves and inertial oscillations. Periodic and unbounded domains are given particular attention in the analysis. The periodic domain is given special emphasis as in such a setting the waves do not radiate to infinity leaving behind the balanced flow. In real geophysical flows such as in the atmosphere or the ocean, waves and balanced flow interact continuously for very long times. A finite domain with periodic boundary conditions acts as the simplest test bed to study such long-time interactions between waves and balanced flow.

In § 2, we proceed systematically using asymptotics to derive a higher-order slow evolution equation in RSW that is valid for $t \lesssim 1/(\epsilon^2 f)$, i.e. a time scale exceeding that of QG dynamics (which is $t \lesssim 1/(\epsilon f)$). This new slow equation, accurate up to $O(\epsilon^2)$, is used to argue that waves can influence the slow dynamics. We then show that the slow energy equation is not closed, but can exchange energy with fast waves. Potential enstrophy, which is seen to be an exact invariant of this new slow equation, is also affected by the waves. In § 3, we set up the problem in a periodic domain and use resonant wave interaction theory to construct a simple reduced system that illustrates fast waves influencing balanced flow. In this special system, the waves act as catalysts, modifying energy exchange between geostrophic modes. Section 4 summarizes the findings of the present work.

2. General results

We will use the method of multiple time scale asymptotics (see, e.g., Ablowitz (2011), Chap. 4 for the particular method that we employ) to investigate the interaction between fast modes and slow balanced flow in RSW. Although the analysis in this section is quite general, we will later place special emphasis on periodic and unbounded domains. In the case of unbounded domains, compact support for initial data is not strictly enforced.

The RSW equations representing a thin layer (small aspect ratio) of fluid on a rotating plane are

$$\left.\begin{aligned} \frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{f}\times\boldsymbol{v} + g\nabla h + \nabla\left(\frac{\boldsymbol{v}^2}{2}\right) + \hat{z}\times\zeta\,\boldsymbol{v} = 0, \\ \frac{\partial h}{\partial t} + \nabla\cdot\{(H+h)\boldsymbol{v}\} = 0, \end{aligned}\right\} \tag{2.1}$$

where $\boldsymbol{f} = f\hat{z}$ is the constant rotation rate ($\hat{z}$ is the unit vector along the z direction), $\zeta\hat{z} = \nabla\times\boldsymbol{v}$ is the vorticity, g is the acceleration due to gravity, H is the constant mean fluid height and h refers to the fluctuations about the mean, which will be assumed to be asymptotically small with respect to H in this work.

We scale variables as $\boldsymbol{x}\to\sqrt{gH}/f\boldsymbol{x}$, $t\to t/f$, $\boldsymbol{v}\to U\boldsymbol{v}$, $h\to U/\sqrt{gH}\,Hh$, where the deformation scale, $\sqrt{gH}/f$, is a natural length scale arising in RSW systems. For this scaling, $\epsilon = Ro = Fr = U/\sqrt{gH}$, where Ro and Fr are the Rossby and Froude numbers respectively and our interest is in the regime $\epsilon \ll 1$. To capture the slowly evolving dynamics, we define a slow time scale, $T=\epsilon t$, and split the time derivative in the above equations as $\partial/\partial t \to \partial/\partial t + \epsilon\partial/\partial T$. The previous equations are then modified as

$$\frac{\partial \boldsymbol{v}}{\partial t} + \hat{z}\times\boldsymbol{v} + \nabla h + \epsilon\left(\frac{\partial \boldsymbol{v}}{\partial T} + \nabla\left(\frac{\boldsymbol{v}^2}{2}\right) + \hat{z}\times\zeta\,\boldsymbol{v}\right) = 0, \tag{2.2a}$$

$$\frac{\partial h}{\partial t} + \nabla\cdot\boldsymbol{v} + \epsilon\left(\frac{\partial h}{\partial T} + \nabla\cdot(h\boldsymbol{v})\right) = 0, \tag{2.2b}$$

$$\frac{\partial q}{\partial t} + \epsilon\left(\frac{\partial q}{\partial T} + \nabla\cdot(q\boldsymbol{v})\right) = 0, \tag{2.2c}$$

where (2.2*c*), the evolution equation for the linear PV, $q = \zeta - h$, was obtained by subtracting (2.2*b*) from the curl of (2.2*a*).

We define averaging over fast time as

$$\overline{\psi}(\boldsymbol{x}, T) = \lim_{\tilde{T}\to\infty}\frac{1}{\tilde{T}}\int_0^{\tilde{T}}\psi(\boldsymbol{x}, t, T)\,\mathrm{d}t, \tag{2.3}$$

so that all variables may be split into a slow and a fast part as $\psi(\boldsymbol{x}, t, T) = \overline{\psi}(\boldsymbol{x}, T) + \psi'(\boldsymbol{x}, t, T)$, such that $\overline{\psi'} = 0$.

2.1. *Asymptotic analysis*

All fields are now expanded in asymptotic series:

$$(\boldsymbol{v}, h, q\ldots) = (\boldsymbol{v}_0, h_0, q_0\ldots) + \epsilon(\boldsymbol{v}_1, h_1, q_1\ldots) + \epsilon^2(\boldsymbol{v}_2, h_2, q_2\ldots) + \cdots. \tag{2.4}$$

O(1) *equations*

The equations at leading order are linear:

$$\frac{\partial \boldsymbol{v}_0}{\partial t} + \hat{\boldsymbol{z}} \times \boldsymbol{v}_0 + \boldsymbol{\nabla} h_0 = 0, \tag{2.5a}$$

$$\frac{\partial h_0}{\partial t} + \boldsymbol{\nabla} \cdot \boldsymbol{v}_0 = 0, \tag{2.5b}$$

$$\frac{\partial q_0}{\partial t} = 0. \tag{2.5c}$$

The solution of the above linear system can be split into three different parts – inertial oscillations, inertio-gravity waves and a geostrophically balanced part (represented by subscripts '*IO*', '*W*' and '*G*' respectively hereafter). We write the solution of the above system as a sum of these different parts:

$$\boldsymbol{v}_0 = \boldsymbol{v}_{IO} + \boldsymbol{v}_W + \boldsymbol{v}_G, \quad h_0 = h_{IO} + h_W + h_G. \tag{2.6a,b}$$

The equations that each of these parts separately satisfy are

$$\left.\begin{aligned} \frac{\partial \boldsymbol{v}_{IO}}{\partial t} + \hat{\boldsymbol{z}} \times \boldsymbol{v}_{IO} = 0, \quad \frac{\partial \boldsymbol{v}_W}{\partial t} + \hat{\boldsymbol{z}} \times \boldsymbol{v}_W + \boldsymbol{\nabla} h_W = 0, \quad \hat{\boldsymbol{z}} \times \boldsymbol{v}_G + \boldsymbol{\nabla} h_G = 0, \\ \frac{\partial h_{IO}}{\partial t} = 0, \quad \frac{\partial h_W}{\partial t} + \boldsymbol{\nabla} \cdot \boldsymbol{v}_W = 0, \quad \boldsymbol{\nabla} \cdot \boldsymbol{v}_G = 0. \end{aligned}\right\} \tag{2.7}$$

Inertial oscillations are spatially uniform and are pure oscillations at inertial frequency. The height field of inertial oscillations is constant and is thus set to zero, i.e. $h_{IO} = 0$, since the constant mean height has already been removed. Inertio-gravity waves have superinertial frequency and form the fastest evolving part of the linear system. On the other hand, the divergence-free geostrophically balanced part does not evolve in this fast time (as may be inferred from (2.5*c*) and (2.8) below) but is expected to evolve on a slower time scale. Of course, the fast waves will also evolve in this slow time scale. From (2.7), it follows that

$$q_W = 0 \Leftrightarrow \zeta_W = h_W \quad \text{and} \tag{2.8a}$$

$$q_0 = q_G = (\Delta - 1) h_G, \tag{2.8b}$$

where $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2$. Equation (2.8*a*) implies that the leading-order PV is unaffected by waves, since the vorticity and the height field associated with the waves are equal. We shall make use of this relationship, i.e. (2.8*a*), in several algebraic manipulations that follow.

Thus, at leading order, a strict splitting exists between balanced and unbalanced fields, with the fast unbalanced part consisting of inertio-gravity waves and inertial oscillations, and the slow part consisting of geostrophic balanced flow. This splitting may be expressed as

$$\boldsymbol{v}_0 = \overline{\boldsymbol{v}}_0(\boldsymbol{x}, T) + \boldsymbol{v}_0'(\boldsymbol{x}, t, T), \quad \overline{\boldsymbol{v}}_0 = \boldsymbol{v}_G \text{ and } \boldsymbol{v}_0' = \boldsymbol{v}_{IO} + \boldsymbol{v}_W, \tag{2.9a}$$

$$h_0 = \overline{h}_0(\boldsymbol{x}, T) + h_0'(\boldsymbol{x}, t, T), \quad \overline{h}_0 = h_G \text{ and } h_0' = h_W, \tag{2.9b}$$

$$q_0 = \overline{q}_0(\boldsymbol{x}, T) + q_0'(\boldsymbol{x}, t, T), \quad \overline{q}_0 = q_G \text{ and } q_0' = 0. \tag{2.9c}$$

$O(\epsilon)$ equations

The equations at this order are

$$\frac{\partial \boldsymbol{v}_1}{\partial t} + \frac{\partial \boldsymbol{v}_0}{\partial T} + \hat{\boldsymbol{z}} \times (\boldsymbol{v}_1 + \zeta_0 \boldsymbol{v}_0) + \boldsymbol{\nabla} \left(h_1 + \frac{\boldsymbol{v}_0^2}{2} \right) = 0, \tag{2.10a}$$

$$\frac{\partial h_1}{\partial t} + \frac{\partial h_0}{\partial T} + \boldsymbol{\nabla} \boldsymbol{\cdot} (\boldsymbol{v}_1 + h_0 \boldsymbol{v}_0) = 0, \tag{2.10b}$$

$$\frac{\partial q_1}{\partial t} + \frac{\partial q_0}{\partial T} + \boldsymbol{\nabla} \boldsymbol{\cdot} (q_0 \boldsymbol{v}_0) = 0. \tag{2.10c}$$

We split the solution of the above system into fast and slow parts as

$$\boldsymbol{v}_1 = \overline{\boldsymbol{v}}_1(\boldsymbol{x}, T) + \boldsymbol{v}_1'(\boldsymbol{x}, t, T), \quad h_1 = \overline{h}_1(\boldsymbol{x}, T) + h_1'(\boldsymbol{x}, t, T), \quad q_1 = \overline{q}_1(\boldsymbol{x}, T) + q_1'(\boldsymbol{x}, t, T), \tag{2.11a–c}$$

such that $\overline{\boldsymbol{v}_1'} = \overline{h_1'} = \overline{q_1'} = 0$. For arbitrary initial data, the initial condition at $O(\epsilon)$ (which may or may not be zero) will be satisfied by the sum of the fast and slow parts of each variable. We then substitute (2.11) into (2.10) and average in fast time to obtain the slow equations:

$$\frac{\partial \boldsymbol{v}_G}{\partial T} + \hat{\boldsymbol{z}} \times (\overline{\boldsymbol{v}}_1 + \zeta_G \boldsymbol{v}_G + \overline{h_W \boldsymbol{v}_W}) + \boldsymbol{\nabla} \left(\overline{h}_1 + \frac{\boldsymbol{v}_G^2}{2} + \frac{\overline{\boldsymbol{v}_W^2}}{2} \right) = 0, \tag{2.12a}$$

$$\frac{\partial h_G}{\partial T} + \boldsymbol{\nabla} \boldsymbol{\cdot} (\overline{\boldsymbol{v}}_1 + h_G \boldsymbol{v}_G + \overline{h_W \boldsymbol{v}_W}) = 0, \tag{2.12b}$$

$$\frac{\partial q_0}{\partial T} + \boldsymbol{\nabla} \boldsymbol{\cdot} (q_0 \boldsymbol{v}_G) = 0. \tag{2.12c}$$

It should be noted that interaction terms of the form IO–IO and IO–W do not appear in the above slow equations due to the spatial homogeneity of IOs. Equation (2.12*c*), which arises as a solvability condition for extending the validity of multiscale asymptotics to $T \sim O(1)$, is the famous QG equation. At this order of asymptotics, the fast modes do not affect the slow evolution of the balanced flow, and as a result one can describe the slow dynamics of the balanced flow without having any information on the fast modes.

We now use (2.12) to obtain the $O(\epsilon)$ slow velocity field. Unlike the case at leading order, (2.9), an exact splitting between balanced and unbalanced modes does not exist anymore and the slow velocity field at $O(\epsilon)$ consists of wave–wave (WW) interaction terms in addition to geostrophic–geostrophic (GG) terms. We will split the slow terms we obtain at this order of asymptotics into two parts – a part that emerges from wave interactions alone, and the remaining part that would be present even in the absence of wave activity. From (2.12*a*) and (2.12*b*) we obtain

$$\overline{\boldsymbol{v}}_1 = \overline{\boldsymbol{v}}_{1W} + \overline{\boldsymbol{v}}_{1G}, \quad \text{where} \tag{2.13a}$$

$$\overline{\boldsymbol{v}}_{1W} = -\overline{h_W \boldsymbol{v}_W} + \hat{\boldsymbol{z}} \times \boldsymbol{\nabla} \left(\frac{\overline{\boldsymbol{v}_W^2}}{2} \right) \quad \text{and} \tag{2.13b}$$

$$\overline{\boldsymbol{v}}_{1G} = -\zeta_G \boldsymbol{v}_G + \hat{\boldsymbol{z}} \times \left\{ \frac{\partial \boldsymbol{v}_G}{\partial T} + \boldsymbol{\nabla} \left(\overline{h}_1 + \frac{\boldsymbol{v}_G^2}{2} \right) \right\}, \tag{2.13c}$$

$$\frac{\partial h_G}{\partial T} = -\boldsymbol{\nabla} \boldsymbol{\cdot} \overline{\boldsymbol{v}}_1, \tag{2.13d}$$

where $\overline{\boldsymbol{v}}_{1W}$ is the fast-time independent part of $\boldsymbol{v}_1$ due to nonlinear wave–wave interactions and $\overline{\boldsymbol{v}}_{1G}$ is the remaining part of the slow velocity field. In the absence of waves, $\overline{\boldsymbol{v}}_1 = \overline{\boldsymbol{v}}_{1G}$. Equation (2.13*d*) follows directly from (2.12*b*) if we observe that $\boldsymbol{\nabla}\boldsymbol{\cdot}(h_G\boldsymbol{v}_G) = \boldsymbol{\nabla}\boldsymbol{\cdot}(\overline{h_W\boldsymbol{v}_W}) = 0$ (using (2.7)). It should also be noted that since (2.12*a*,*b*) are not linearly independent (but are related via (2.12*c*)), (2.13*d*) is not a separate condition but may be obtained by taking the divergence of (2.13*a*) and then using (2.12*c*). From (2.13*a*) we obtain

$$\overline{q}_1 = \overline{\zeta}_1 - \overline{h}_1 = \hat{\boldsymbol{z}}\boldsymbol{\cdot}(\boldsymbol{\nabla}\times\overline{\boldsymbol{v}}_1) - \overline{h}_1 = \overline{q}_{1W} + \overline{q}_{1G}, \quad \text{where} \tag{2.14a}$$

$$\overline{q}_{1W} = \overline{\zeta}_{1W} = \hat{\boldsymbol{z}}\boldsymbol{\cdot}(\boldsymbol{\nabla}\times\overline{\boldsymbol{v}}_{1W}) = \varDelta\frac{\overline{\boldsymbol{v}_W^2}}{2} - \overline{h_W^2} - \hat{\boldsymbol{z}}\boldsymbol{\cdot}(\overline{\boldsymbol{\nabla}h_W\times\boldsymbol{v}_W}) \quad \text{and} \tag{2.14b}$$

$$\overline{q}_{1G} = \overline{\zeta}_{1G} - \overline{h}_1 = \hat{\boldsymbol{z}}\boldsymbol{\cdot}(\boldsymbol{\nabla}\times\overline{\boldsymbol{v}}_{1G}) - \overline{h}_1 = (\varDelta - 1)\overline{h}_1 + \varDelta\frac{\boldsymbol{v}_G^2}{2} - \boldsymbol{\nabla}\boldsymbol{\cdot}(\zeta_G\boldsymbol{\nabla}h_G). \tag{2.14c}$$

Here, similarly to the velocity splitting, we split the slow PV at $O(\epsilon)$ into two parts, identifying a part that depends only on wave interactions. If the initial data were completely balanced, $\overline{q}_1 = \overline{q}_{1G}$. Subtracting (2.12*c*) from (2.10*c*) gives the fast PV equation at $O(\epsilon)$:

$$\frac{\partial q_1'}{\partial t} = -\boldsymbol{\nabla}\boldsymbol{\cdot}\{q_0(\boldsymbol{v}_{IO}+\boldsymbol{v}_W)\} = q_0\frac{\partial h_W}{\partial t} - \boldsymbol{v}_W\boldsymbol{\cdot}\boldsymbol{\nabla}q_0 - \boldsymbol{v}_{IO}\boldsymbol{\cdot}\boldsymbol{\nabla}q_0, \tag{2.15}$$

which can be integrated to obtain

$$q_1' = q_{1W}' + q_{1IO}', \quad \text{where} \tag{2.16a}$$

$$q_{1W}' = q_0h_W - \boldsymbol{\nabla}q_0\boldsymbol{\cdot}\int^t \boldsymbol{v}_W\,\mathrm{d}t \quad \text{and} \tag{2.16b}$$

$$q_{1IO}' = -\boldsymbol{\nabla}q_0\boldsymbol{\cdot}\int^t \boldsymbol{v}_{IO}\,\mathrm{d}t. \tag{2.16c}$$

Here, $\int^t$ stands for evaluating the integrated function at t. (We note that the expressions in (2.16) can be expressed in alternate forms. For instance, from (2.7), one can show that $(\partial^2/\partial t^2 + 1 - \varDelta)\boldsymbol{v}_W = 0$, which may be used to obtain $\boldsymbol{\nabla}q_0\boldsymbol{\cdot}\int^t \boldsymbol{v}_W\,\mathrm{d}t = \{\boldsymbol{\nabla}q_0\boldsymbol{\cdot}[(\varDelta-1)^{-1}\boldsymbol{v}_W]\}_t$. Similarly, $\boldsymbol{\nabla}q_0\boldsymbol{\cdot}\int^t \boldsymbol{v}_{IO}\,\mathrm{d}t = -\{\boldsymbol{\nabla}q_0\boldsymbol{\cdot}\boldsymbol{v}_{IO}\}_t$. From these alternate expressions, it is easier to see that $\overline{q_{1W}'} = \overline{q_{1IO}'} = 0$. An equivalent alternative approach is to use the primitive of fast variables, as in RZB (see their equations (3.35) and (3.33)).)

Thus, at this point we have obtained the full PV at $O(\epsilon)$, given by the sum of (2.14*a*) and (2.16*a*), and the fast-time independent part of the $O(\epsilon)$ velocity field given by (2.13*a*). To obtain a higher-order slow evolution equation of RSW valid for $t \lesssim 1/\epsilon^2$, we rewrite (2.12*c*) as

$$\frac{\partial q_0}{\partial T} + \boldsymbol{\nabla}\boldsymbol{\cdot}(q_0\boldsymbol{v}_G) = \epsilon\varPhi(\boldsymbol{x}, T) + O(\epsilon^2). \tag{2.17}$$

We will obtain the correction term $\varPhi$ by eliminating resonance at the next order of asymptotics ($O(\epsilon^2)$). We also retain an '$O(\epsilon^2)$' on the right-hand side of the above equation (a convention that we shall follow hereafter) to emphasis that this higher-order equation that we derive holds for $T \lesssim 1/\epsilon$, but not for $T \gg 1/\epsilon$. The reader is referred to appendix A for more details related to this technique of obtaining $O(\epsilon)$ correction terms by preventing resonance at $O(\epsilon^2)$ and thus obtaining an equation valid for time scales $t \lesssim 1/\epsilon^2$, and also for a comparison with multiscale asymptotics using two slow time scales.

$O(\epsilon^2)$ equations

The PV equation at this order is

$$\frac{\partial q_2}{\partial t}+\Phi(\boldsymbol{x},T)+\frac{\partial q_1}{\partial T}+\boldsymbol{\nabla}\boldsymbol{\cdot}(q_0\boldsymbol{v}_1+q_1\boldsymbol{v}_0)=0. \tag{2.18}$$

Substituting (2.9) and (2.11) into (2.18) and averaging over fast time leads to

$$\Phi(\boldsymbol{x},T)+\frac{\partial \overline{q}_1}{\partial T}+\boldsymbol{\nabla}\boldsymbol{\cdot}(q_0\overline{\boldsymbol{v}}_1+\overline{q}_1\boldsymbol{v}_G+\overline{q'_{1W}\boldsymbol{v}_W}+\overline{q'_{1IO}\boldsymbol{v}_{IO}})=0, \tag{2.19}$$

where, once again, we note that interaction terms involving inertial oscillations and inertio-gravity waves do not project onto the above slow equation. Furthermore, the interaction term above due to inertial oscillations remarkably cancels, as shown below:

$$\begin{aligned}\boldsymbol{\nabla}\boldsymbol{\cdot}\overline{(q'_{1IO}\boldsymbol{v}_{IO})} &= \overline{\boldsymbol{v}_{IO}\boldsymbol{\cdot}\boldsymbol{\nabla}q'_{1IO}}=-\overline{(\boldsymbol{v}_{IO}\boldsymbol{\cdot}\boldsymbol{\nabla})\left(\int^t \mathrm{d}t(\boldsymbol{v}_{IO}\boldsymbol{\cdot}\boldsymbol{\nabla})\right)q_0}\\ &= -\overline{\left(\left(\int^t \mathrm{d}t(\boldsymbol{v}_{IO}\boldsymbol{\cdot}\boldsymbol{\nabla})\right)^2\frac{q_0}{2}\right)_t}=0.\end{aligned} \tag{2.20}$$

Thus, inertial oscillations do not affect the balanced flow up to $O(\epsilon^2)$.

Using (2.19) and (2.20) in (2.17) gives

$$\frac{\partial}{\partial T}(q_0+\epsilon\overline{q}_1)+\boldsymbol{\nabla}\boldsymbol{\cdot}\{(q_0+\epsilon\overline{q}_1)\boldsymbol{v}_G+\epsilon q_0\overline{\boldsymbol{v}}_1+\epsilon\overline{q'_{1W}\boldsymbol{v}_W}\}+O(\epsilon^2)=0, \tag{2.21}$$

which can be written using (2.13*a*) and (2.14*a*) as

$$\begin{aligned}&\frac{\partial}{\partial T}(q_0+\epsilon\overline{q}_{1G}+\epsilon\overline{q}_{1W})\\ &\quad+\boldsymbol{\nabla}\boldsymbol{\cdot}\{(q_0+\epsilon\overline{q}_{1G}+\epsilon\overline{q}_{1W})\boldsymbol{v}_G+\epsilon q_0\overline{\boldsymbol{v}}_{1G}+\epsilon q_0\overline{\boldsymbol{v}}_{1W}+\epsilon\overline{q'_{1W}\boldsymbol{v}_W}\}+O(\epsilon^2)=0.\end{aligned} \tag{2.22}$$

Equation (2.22) is the slow evolution equation of RSW (2.2) for arbitrary initial data up to $O(\epsilon^2)$. With the $O(\epsilon)$ correction terms, this equation describes the evolution of the slow dynamics of RSW for arbitrary and unbalanced initial data for $t\lesssim 1/\epsilon^2$, (2.12*c*) being the slow evolution equation for time scales $t\lesssim 1/\epsilon$. One can also express the above equation in an expanded form using the $O(\epsilon)$ fields that we obtained before. For instance, substituting (2.13*b*,*c*), (2.14*b*,*c*) and (2.16*b*) into (2.22), we obtain

$$\begin{aligned}&\frac{\partial}{\partial T}\left\{(\Delta-1)(h_G+\epsilon\overline{h}_1)+\epsilon\left(\Delta\frac{\boldsymbol{v}_G^2}{2}-\boldsymbol{\nabla}\boldsymbol{\cdot}(\zeta_G\boldsymbol{\nabla}h_G)+\Delta\frac{\overline{\boldsymbol{v}_W^2}}{2}-\overline{h_W^2}-\hat{\boldsymbol{z}}\boldsymbol{\cdot}(\overline{\boldsymbol{\nabla}h_W\times\boldsymbol{v}_W})\right)\right\}\\ &\quad+\boldsymbol{\nabla}\boldsymbol{\cdot}\left[\left\{(\Delta-1)(h_G+\epsilon\overline{h}_1)\right.\right.\\ &\quad+\left.\epsilon\left(\Delta\frac{\boldsymbol{v}_G^2}{2}-\boldsymbol{\nabla}\boldsymbol{\cdot}(\zeta_G\boldsymbol{\nabla}h_G)+\Delta\frac{\overline{\boldsymbol{v}_W^2}}{2}-\overline{h_W^2}-\hat{\boldsymbol{z}}\boldsymbol{\cdot}(\overline{\boldsymbol{\nabla}h_W\times\boldsymbol{v}_W})\right)\right\}\boldsymbol{v}_G\\ &\quad+\epsilon q_0\left\{-\zeta_G\boldsymbol{v}_G+\hat{\boldsymbol{z}}\times\left\{\frac{\partial\boldsymbol{v}_G}{\partial T}+\boldsymbol{\nabla}\left(\overline{h}_1+\frac{\boldsymbol{v}_G^2}{2}\right)\right\}\right\}\end{aligned}$$

$$+\epsilon\hat{z}\times q_0\nabla\left(\frac{\boldsymbol{v}_W^2}{2}\right)-\epsilon\overline{\left(\nabla q_0\cdot\int^t \boldsymbol{v}_W\,\mathrm{d}t\right)\boldsymbol{v}_W}\Bigg]$$
$$+O(\epsilon^2)=0. \tag{2.23}$$

Equation (2.23) which describes the slow evolution of the total slow height field $h_G+\epsilon\overline{h}_1$ (or $\overline{h}_0+\epsilon\overline{h}_1$) is influenced by the fast waves and hence by itself does not form a closed equation. An equivalent slow evolution equation for the waves along with (2.23) will form a coupled system that describes the fast–slow interactions in RSW for very long time scales.

In the absence of waves, one simply sets the wave interaction terms in (2.22) to zero to obtain the slow evolution of balanced data:

$$\frac{\partial}{\partial T}(q_0+\epsilon\overline{q}_{1G})+\nabla\cdot\{(q_0+\epsilon\overline{q}_{1G})\boldsymbol{v}_G+\epsilon q_0\overline{\boldsymbol{v}}_{1G}\}+O(\epsilon^2)=0, \tag{2.24}$$

or in expanded form (which follows directly from (2.23))

$$\frac{\partial}{\partial T}\left\{(\Delta-1)(h_G+\epsilon\overline{h}_1)+\epsilon\left(\Delta\frac{\boldsymbol{v}_G^2}{2}-\nabla\cdot(\zeta_G\nabla h_G)\right)\right\}$$
$$+\nabla\cdot\left[\left\{(\Delta-1)(h_G+\epsilon\overline{h}_1)+\epsilon\left(\Delta\frac{\boldsymbol{v}_G^2}{2}-\nabla\cdot(\zeta_G\nabla h_G)\right)\right\}\boldsymbol{v}_G\right.$$
$$\left.+\epsilon q_0\left\{-\zeta_G\boldsymbol{v}_G+\hat{z}\times\left\{\frac{\partial\boldsymbol{v}_G}{\partial T}+\nabla\left(\overline{h}_1+\frac{\boldsymbol{v}_G^2}{2}\right)\right\}\right\}\right]+O(\epsilon^2)=0. \tag{2.25}$$

The slow evolution equation (2.24) or (2.25), unlike (2.23), forms a closed system since it is unaffected by fast waves. This higher-order balance equation, (2.12*c*) being the lower-order one, is in the same spirit as quasi-geostrophy since it describes the evolution of the balanced slow height field $(h_G+\epsilon\overline{h}_1)$ by completely ignoring the fast waves (note that knowledge of h_G fixes all other geostrophic terms in (2.25) using the linear relationships (2.7) and (2.8*b*)). With initial conditions for $h_G+\epsilon\overline{h}_1$, one may numerically integrate (2.25) (or equivalent forms – see appendix B) to track the evolution of slow balanced flow ignoring fast waves. Furthermore, (2.24) conserves energy and enstrophy, as we show below.

Although a generalized higher-order slow equation for the evolution of arbitrary initial data in RSW (such as (2.22) or (2.23)) does not seem to have been derived before, several authors (Allen 1993, Warn *et al.* 1995, RZB) have obtained equivalent forms of the higher-order balance equation (2.25) in different settings. For instance, while Warn *et al.* (1995) ignore fast waves and use slaving principles to obtain the higher-order balanced equation, RZB admit arbitrary but compact supported initial data that lead to the dispersive decay of wave fields, leading to geostrophic adjustment. Since our approach is closest to that of RZB, using multi-time-scale asymptotics, we show in appendix B that our slow equation devoid of wave activity, (2.24), is equivalent to the higher-order balance equation of RZB.

We conclude the derivation of higher-order slow equations by emphasizing that if the initial data are arbitrary and can excite fast waves, the slow dynamics is affected by fast waves, and thus the slow evolution equation is not closed in the long-time limit. This breaks down the basic idea of quasi-geostrophy since to track the slow dynamics, one needs to know the parallel evolution of fast dynamics, due to the fast–slow coupling.

The special case of an unbounded domain with compact initial data

It is interesting to note that in the case of an unbounded domain, if initial data have compact support, the interaction terms due to waves identically vanish. This is because in such a set-up, the amplitude of the waves decays as $1/t$, as demonstrated by RZB using the standard stationary-phase calculation. For instance, in (2.13*b*),

$$\overline{h_W \boldsymbol{v}_W} \sim \frac{\overline{\boldsymbol{v}_W^2}}{2} \sim \lim_{\tilde{T}\to\infty} \frac{1}{\tilde{T}} \int^{\tilde{T}} \frac{c}{t^2}\,\mathrm{d}t = 0 \Rightarrow \overline{\boldsymbol{v}}_{1W} = 0. \tag{2.26}$$

Other wave–wave terms also vanish in similar fashion on averaging, and as a result the slow evolution equation is devoid of wave activity. Thus, although the initial data are arbitrary and thus unbalanced, the balanced flow remains unaffected by the waves up to $O(\epsilon^2)$, and an adjustment to a balanced state would always exist in such a set-up. Due to the algebraic decay of wave amplitudes, RZB mention the possibility of a lack of fast modes affecting the balanced flow even in higher-order correction terms.

However, in a generic domain, the wave interaction terms are non-zero and thus the waves will affect the slow dynamics of the balanced flow. As a consequence, solving for the slow dynamics of the balanced flow would require knowledge of the wave dynamics.

2.2. *Energy and enstrophy exchange between waves and balanced flow*

Energy and enstrophy of RSW

We now investigate the possibility of energy and enstrophy exchange between fast and slow dynamics. We begin by recalling the energy equation of RSW (2.2),

$$\frac{\mathrm{d}}{\mathrm{d}T}\overline{\left\langle (1+\epsilon h)\frac{\boldsymbol{v}^2}{2} + \frac{h^2}{2}\right\rangle} = 0, \tag{2.27}$$

where we averaged over the fast time t and $\langle\ \ \rangle$ refers to the spatial averaging operation. For example, in the case of a periodic domain $[-L_x, L_x]\times[-L_y, L_y]$,

$$\langle F\rangle = \frac{1}{4L_xL_y}\int_{-L_x}^{L_x}\int_{-L_y}^{L_y} F\,\mathrm{d}x\,\mathrm{d}y, \tag{2.28}$$

whereas the limit $L_x, L_y \to \infty$ would be taken in the case of an unbounded domain. We substitute the asymptotic expansions (2.4) in (2.27) to obtain the energy conservation equation of RSW up to $O(\epsilon^2)$,

$$\frac{\mathrm{d}}{\mathrm{d}T}\overline{\left\langle (1+\epsilon h_0)\left(\frac{\boldsymbol{v}_0^2}{2}\right) + \epsilon \boldsymbol{v}_0\boldsymbol{\cdot}\boldsymbol{v}_1 + \left(\frac{h_0^2}{2} + \epsilon h_0 h_1\right)\right\rangle} + O(\epsilon^2) = 0. \tag{2.29}$$

Another integral invariant associated with RSW is the Ertel potential enstrophy. This Casimir corresponding to (2.2) is given by (after fast time averaging)

$$\frac{\mathrm{d}}{\mathrm{d}T}\overline{\left\langle \frac{(1+\epsilon\zeta)^2}{(1+\epsilon h)}\right\rangle} = 0. \tag{2.30}$$

Now,

$$\frac{(1+\epsilon\zeta)^2}{(1+\epsilon h)} = \frac{\{(1+\epsilon h)+\epsilon q\}^2}{(1+\epsilon h)} = 1 + \epsilon(h+2q) + \epsilon^2\frac{q^2}{1+\epsilon h}, \tag{2.31}$$

where we used $\zeta = q + h$. Substituting (2.31) in (2.30) after simplification gives

$$\frac{\mathrm{d}}{\mathrm{d}T}\overline{\langle q^2(1-\epsilon h)\rangle} + O(\epsilon^2) = 0. \tag{2.32}$$

Substituting (2.4) in (2.32) gives the enstrophy equation up to $O(\epsilon^2)$,

$$\frac{\mathrm{d}}{\mathrm{d}T}\overline{\langle q_0^2(1-\epsilon h_0) + 2\epsilon q_0 q_1\rangle} + O(\epsilon^2) = 0. \tag{2.33}$$

The slow energy equation

After multiplying (2.22) by h_G and some rearrangements, we obtain

$$\begin{aligned} & h_G\frac{\partial q_0}{\partial T} + \epsilon\frac{\partial}{\partial T}\{h_G(\overline{q}_{1G}+\overline{q}_{1W})\} \\ & + \nabla\cdot\{h_G((q_0+\epsilon\overline{q}_{1G}+\epsilon\overline{q}_{1W})\boldsymbol{v}_G + \epsilon q_0(\overline{\boldsymbol{v}}_{1W}+\overline{\boldsymbol{v}}_{1G}) + \epsilon\overline{q'_{1W}\boldsymbol{v}_W})\} \\ & - \epsilon\left\{\overline{q}_{1G}\frac{\partial h_G}{\partial T} + q_0\nabla h_G\cdot\overline{\boldsymbol{v}}_{1G} + \overline{q}_{1W}\frac{\partial h_G}{\partial T} + q_0\nabla h_G\cdot\overline{\boldsymbol{v}}_{1W} + \overline{(\nabla h_G\cdot\boldsymbol{v}_W)q'_{1W}}\right\} \\ & + O(\epsilon^2) = 0. \end{aligned} \tag{2.34}$$

After some straightforward manipulations, we obtain the energy equation corresponding to the slow dynamics (the reader may refer to appendix C for the details involved in the derivation),

$$\begin{aligned} & \frac{\mathrm{d}}{\mathrm{d}T}\left\langle (1+\epsilon h_G)\frac{\boldsymbol{v}_G^2}{2} + \frac{h_G^2}{2} + \epsilon(\boldsymbol{v}_G\cdot\overline{\boldsymbol{v}}_{1G} + h_G\overline{h}_1)\right\rangle = -\epsilon\frac{\mathrm{d}}{\mathrm{d}T}\langle\boldsymbol{v}_G\cdot\overline{\boldsymbol{v}}_{1W}\rangle \\ & \quad - \epsilon\left\langle\left(\frac{\overline{\boldsymbol{v}_W^2}}{2}\right)\frac{\partial h_G}{\partial T} + \frac{\partial\boldsymbol{v}_G}{\partial T}\cdot\overline{h_W\boldsymbol{v}_W}\right\rangle + \epsilon\overline{\left\langle(\nabla h_G\cdot\boldsymbol{v}_W)\left(\nabla q_0\cdot\int^t\boldsymbol{v}_W\,\mathrm{d}t\right)\right\rangle} \\ & \quad + O(\epsilon^2) = 0. \end{aligned} \tag{2.35}$$

It is seen that the energy associated with the slow equation is not conserved, but contains terms that are potentially capable of energy exchange with fast dynamics. To see this more explicitly, we substitute (2.6) (ignoring inertial oscillations) and (2.11) in (2.29) to obtain the complete energy conservation equation of RSW up to $O(\epsilon^2)$,

$$\begin{aligned} & \frac{\mathrm{d}}{\mathrm{d}T}\left\langle (1+\epsilon h_G)\frac{\boldsymbol{v}_G^2}{2} + \frac{h_G^2}{2} + \epsilon(\boldsymbol{v}_G\cdot\overline{\boldsymbol{v}}_{1G} + h_G\overline{h}_1)\right. \\ & \quad \left. + (1+\epsilon h_G)\frac{\overline{\boldsymbol{v}_W^2}}{2} + \frac{\overline{h_W^2}}{2} + \epsilon\left(\overline{h_W\boldsymbol{v}_W}\cdot\boldsymbol{v}_G + \boldsymbol{v}_G\cdot\overline{\boldsymbol{v}}_{1W} + \overline{\boldsymbol{v}_W\cdot\boldsymbol{v}'_1} + \overline{h_W h'_1} + \overline{h_W\frac{\boldsymbol{v}_W^2}{2}}\right)\right\rangle \\ & \quad + O(\epsilon^2) = 0. \end{aligned} \tag{2.36}$$

On comparing (2.35) and (2.36) we see that only part of the total energy is associated with the slow equation, which itself is not conserved, i.e. it may be exchanged with the fast dynamics (in particular, note the appearance of terms such as $\mathrm{d}\langle h_G\overline{\boldsymbol{v}_W^2}/2\rangle/\mathrm{d}T$ in the full energy conservation equation, (2.36), while part of the term, $\langle(\partial h_G/\partial T)\overline{\boldsymbol{v}_W^2}/2\rangle$, appears in the slow energy equation, (2.35)). A long-time evolution equation for the fast waves and the corresponding energy equation associated with the fast dynamics would in combination with the slow energy equation, (2.35), close the energy budget, yielding the total energy conservation equation, (2.36). Thus, energetic interaction between the fast and slow dynamics is possible in the long-time limit. We also note that in the absence of waves, the slow evolution equation associated with balanced dynamics, (2.24), exactly conserves energy, given by

$$\frac{\mathrm{d}}{\mathrm{d}T}\left\langle (1+\epsilon h_G)\frac{\boldsymbol{v}_G^2}{2}+\frac{h_G^2}{2}+\epsilon(\boldsymbol{v}_G\boldsymbol{\cdot}\overline{\boldsymbol{v}}_{1G}+h_G\overline{h}_1)\right\rangle + O(\epsilon^2)=0, \tag{2.37}$$

which may be obtained by setting all wave interaction terms in (2.35) (or (2.36)) to zero. We therefore conclude that the energetics associated with the balanced flow is different in the presence and absence of waves.

The slow enstrophy equation

To obtain the potential enstrophy equation corresponding to the slow dynamics, we multiply (2.21) by q_0 and after some manipulations we obtain

$$\begin{aligned}
&\frac{\partial}{\partial T}(q_0^2+\epsilon q_0\overline{q}_1)-\frac{\partial q_0}{\partial T}(q_0+\epsilon\overline{q}_1)+\boldsymbol{\nabla}\boldsymbol{\cdot}\left\{q_0\left((q_0+\epsilon\overline{q}_1)\boldsymbol{v}_G+\epsilon q_0\overline{\boldsymbol{v}}_1+\epsilon\overline{q'_{1W}\boldsymbol{v}_W}\right)\right\}\\
&\quad-(\boldsymbol{v}_G\boldsymbol{\cdot}\boldsymbol{\nabla}q_0)(q_0+\epsilon\overline{q}_1)-\epsilon\boldsymbol{\nabla}\left(\frac{q_0^2}{2}\right)\boldsymbol{\cdot}\overline{\boldsymbol{v}}_1-\epsilon\boldsymbol{\nabla}q_0\boldsymbol{\cdot}\overline{q'_{1W}\boldsymbol{v}_W}=O(\epsilon^2)\\
&\quad\Rightarrow\frac{\partial}{\partial T}\left(\frac{q_0^2}{2}+\epsilon q_0\overline{q}_1\right)-\epsilon\left(\frac{\partial q_0}{\partial T}+\boldsymbol{v}_G\boldsymbol{\cdot}\boldsymbol{\nabla}q_0\right)\overline{q}_1-\boldsymbol{\nabla}\boldsymbol{\cdot}\left(\frac{q_0^2}{2}\boldsymbol{v}_G\right)\\
&\quad-\epsilon\boldsymbol{\nabla}\boldsymbol{\cdot}\left(\frac{q_0^2}{2}\overline{\boldsymbol{v}}_1\right)+\epsilon\frac{q_0^2}{2}\boldsymbol{\nabla}\boldsymbol{\cdot}\overline{\boldsymbol{v}}_1-\epsilon\boldsymbol{\nabla}q_0\boldsymbol{\cdot}\overline{q'_{1W}\boldsymbol{v}_W}\\
&\quad+\boldsymbol{\nabla}\boldsymbol{\cdot}\{q_0((q_0+\epsilon\overline{q}_1)\boldsymbol{v}_G+\epsilon q_0\overline{\boldsymbol{v}}_1+\epsilon\overline{q'_{1W}\boldsymbol{v}_W})\}=O(\epsilon^2).
\end{aligned} \tag{2.38}$$

We further observe that

$$\begin{aligned}
\boldsymbol{\nabla}q_0\boldsymbol{\cdot}\overline{q'_{1W}\boldsymbol{v}_W}&=\boldsymbol{\nabla}q_0\boldsymbol{\cdot}\left\{q_0\overline{h_W\boldsymbol{v}_W}-\overline{\left(\boldsymbol{\nabla}q_0\boldsymbol{\cdot}\int^t\boldsymbol{v}_W\,\mathrm{d}t\right)\boldsymbol{v}_W}\right\}\\
&=\boldsymbol{\nabla}\boldsymbol{\cdot}\left(\frac{q_0^2}{2}\overline{h_W\boldsymbol{v}_W}\right)-\frac{1}{2}\overline{\left(\left(\int^t\boldsymbol{\nabla}q_0\boldsymbol{\cdot}\boldsymbol{v}_W\,\mathrm{d}t\right)^2\right)_t}=\boldsymbol{\nabla}\boldsymbol{\cdot}\left(\frac{q_0^2}{2}\overline{h_W\boldsymbol{v}_W}\right),
\end{aligned} \tag{2.39}$$

where we used $\boldsymbol{\nabla}\boldsymbol{\cdot}\overline{(h_W\boldsymbol{v}_W)}=0$. Using (2.39), (2.17) and (2.13*d*) in (2.38) and spatially averaging gives

$$\frac{\mathrm{d}}{\mathrm{d}T}\left\langle\frac{q_0^2}{2}+\epsilon q_0\overline{q}_1\right\rangle-\epsilon\left\langle\frac{q_0^2}{2}\frac{\partial h_G}{\partial T}\right\rangle+O(\epsilon^2)=0. \tag{2.40}$$

From (2.17),

$$\begin{aligned}
&h_G\frac{\partial}{\partial T}\left(\frac{q_0^2}{2}\right)+\boldsymbol{\nabla}\boldsymbol{\cdot}\left(\frac{q_0^2}{2}h_G\boldsymbol{v}_G\right)=O(\epsilon)\\
&\Rightarrow\epsilon\left\langle h_G\frac{\partial}{\partial T}\left(\frac{q_0^2}{2}\right)\right\rangle=O(\epsilon^2).
\end{aligned} \tag{2.41}$$

Combining (2.40) and (2.41) gives

$$\frac{\mathrm{d}}{\mathrm{d}T}\left\langle \frac{q_0^2}{2}(1-\epsilon h_G)+\epsilon q_0\overline{q}_1\right\rangle + O(\epsilon^2)=0. \tag{2.42}$$

Thus, $(q_0^2/2)(1-\epsilon h_G)+\epsilon q_0\overline{q}_1$ is an exact integral invariant of (2.21). It should be noted that this invariant is the Ertel potential enstrophy of RSW accurate up to $O(\epsilon^2)$, as we obtained in (2.33). Using (2.14*b*,*c*), we rewrite (2.42) as

$$\frac{\mathrm{d}}{\mathrm{d}T}\left\langle \frac{q_0^2}{2}(1-\epsilon h_G)+\epsilon q_0\overline{q}_{1G}\right\rangle = -\epsilon\frac{\mathrm{d}}{\mathrm{d}T}\langle q_0\overline{q}_{1W}\rangle + O(\epsilon^2)$$

$$\Rightarrow \frac{\mathrm{d}}{\mathrm{d}T}\left\langle \frac{q_0^2}{2}(1-\epsilon h_G)+\epsilon q_0\left((\varDelta-1)\overline{h}_1+\varDelta\frac{\boldsymbol{v}_G^2}{2}-\boldsymbol{\nabla}\boldsymbol{\cdot}(\zeta_G\boldsymbol{\nabla}h_G)\right)\right\rangle \tag{2.43}$$

$$= -\epsilon\frac{\mathrm{d}}{\mathrm{d}T}\left\langle q_0\left(\varDelta\frac{\overline{\boldsymbol{v}_W^2}}{2}-\overline{h_W^2}-\hat{\boldsymbol{z}}\boldsymbol{\cdot}(\overline{\boldsymbol{\nabla}h_W\times\boldsymbol{v}_W})\right)\right\rangle + O(\epsilon^2). \tag{2.44}$$

Thus, although the potential enstrophy is an integral invariant of the slow evolution equation that we derived, it consists of contributions from waves and balanced dynamics. Hence, only part of the total enstrophy will be obtained if we just track the slow evolution of the balanced flow, ignoring fast waves.

The analysis in this section leads to the conclusion that the long-time evolution of slow dynamics in RSW is affected by fast waves, or, in other words, prediction of slow dynamics requires knowledge of the evolution of fast waves. If unbalanced initial data are allowed to evolve in RSW, an adjustment towards a balanced state cannot take place in a generic domain where wave activity does not diminish due to dispersive propagation or dissipation. As a result, QG theory, by which one can predict the slow evolution of the balanced flow only by using the balanced part of the initial data, fails, since knowledge of the fast fields becomes equally important.

If waves are absent or artificially filtered, a balanced flow evolves, as given by the higher-order balance equation. This equation conserves energy and enstrophy and thus forms a closed system. In the presence of waves, however, the slow equation is affected by the fast dynamics, and so are the energy and enstrophy associated with balanced dynamics. Thus, the dynamics of the balanced flow is expected to differ in the presence and absence of fast waves.

Although the interaction terms we derived are of $O(\epsilon)$ for $T\sim O(1)$, significant interactions and energy and enstrophy exchanges between fast waves and the balanced flow are possible in the long-time limit as $T\gg O(1)$. To demonstrate that considerable fast–slow interactions can take place in the long-time limit, we proceed to construct a simple reduced system as a special example in the next section.

3. Quadruple wave-balanced flow interactions in a periodic domain

Our primary goal in this section is to construct a simple example for waves influencing balanced modes in the long-time limit in a doubly periodic domain. Rather than continuing the general analysis in the previous section, we focus on constructing a set of amplitude equations that capture slow–fast interactions using resonant wave interaction theory (Phillips 1960; Benney 1962; Craik 1985). From among the different kinds of wave-balanced flow interactions contained in (2.23), we ignore triad interactions and focus on quadruple wave-balanced mode interactions.

Due to prevalent triad interactions of the form GGG (leading to the QG equation) and WGW (resulting in scattering of waves by G modes), quadruple fast–slow interactions have received very little attention in RSW. This also acts as a strong motive to construct a simple set of amplitude equations that capture waves influencing balanced flow.

3.1. *A reduced system consisting of two waves and two geostrophic modes*

To construct the minimal system, we consider two balanced modes and two wave modes of the form

$$(\boldsymbol{v}_G, h_G) = (\boldsymbol{v}^{G1}, h^{G1}) \exp\{\mathrm{i}\boldsymbol{k}^{G1}\boldsymbol{\cdot}\boldsymbol{x}\} + (\boldsymbol{v}^{G2}, h^{G2}) \exp\{\mathrm{i}\boldsymbol{k}^{G2}\boldsymbol{\cdot}\boldsymbol{x}\} + \text{c.c.}, \tag{3.1a}$$

$$(\boldsymbol{v}_W, h_W) = (\boldsymbol{v}^{W1}, h^{W1}) \exp\{\mathrm{i}(\boldsymbol{k}^{W1}\boldsymbol{\cdot}\boldsymbol{x} - \omega t)\} + (\boldsymbol{v}^{W2}, h^{W2}) \exp\{\mathrm{i}(\boldsymbol{k}^{W2}\boldsymbol{\cdot}\boldsymbol{x} + \omega t)\} + \text{c.c.}, \tag{3.1b}$$

$$\text{such that } |\boldsymbol{k}^{W1}| = |\boldsymbol{k}^{W2}| = k, \quad \omega = \sqrt{1+k^2}, \quad \boldsymbol{k}^{G2} = \boldsymbol{k}^{G1} + \boldsymbol{k}^{W1} + \boldsymbol{k}^{W2}$$
$$\text{and c.c. refers to complex conjugate.} \tag{3.1c}$$

The above system is the minimal set of modes that can be used to derive a quadruple wave-balanced flow system in which triad interactions are absent. This is so because for a single wave mode, all the wave–wave terms affecting the balanced flow exactly cancel. For example, for the above two wave modes, (3.1*b*), from (2.14*b*) we obtain

$$\begin{aligned}\overline{q}_{1W} = & -\big\{(\boldsymbol{k}^{W2}\boldsymbol{\cdot}\boldsymbol{v}^{W1})(\boldsymbol{k}^{W1}+\boldsymbol{k}^{W2})\boldsymbol{\cdot}\boldsymbol{v}^{W2} \\ & + (\boldsymbol{k}^{W1}\boldsymbol{\cdot}\boldsymbol{v}^{W2})(\boldsymbol{k}^{W1}+\boldsymbol{k}^{W2})\boldsymbol{\cdot}\boldsymbol{v}^{W1}\big\} \exp\{\mathrm{i}(\boldsymbol{k}^{W1}+\boldsymbol{k}^{W2})\boldsymbol{\cdot}\boldsymbol{x}\} + \text{c.c.}\end{aligned} \tag{3.2}$$

Modes in (3.1*b*) become a single wave if we set $\boldsymbol{k}^{W2} = -\boldsymbol{k}^{W1}$ and $(\boldsymbol{v}^{W2}, h^{W2}) = (\boldsymbol{v}^{W1^*}, h^{W2^*})$. For this special case of a single wave, $\overline{q}_{1W} = 0$, as may be inferred from (3.2). Similarly, one can show that all wave–wave interaction terms in (2.23) cancel out for the special case of a single wave.

If we imagine a hypothetical situation consisting of only two geostrophic modes, as in (3.1), there are no triad interactions. For such a special setting, we may ignore the term $\boldsymbol{\nabla}\boldsymbol{\cdot}(q_0\boldsymbol{v}_G)$ and redefine the relevant slow time as $\tau = \epsilon T = \epsilon^2 t$. To see the effect of waves on the balanced flow, we also temporarily ignore the interactions between geostrophic modes to rewrite (2.22) in the form

$$\frac{\partial q_0}{\partial \tau} + \boldsymbol{\nabla}\boldsymbol{\cdot}\{\overline{q}_{1W}\boldsymbol{v}_G + q_0\overline{\boldsymbol{v}}_{1W} + \overline{q'_{1W}\boldsymbol{v}_W}\} = 0. \tag{3.3}$$

It should be noted that only the time derivative of leading PV appears above since terms such as $\partial\overline{q}_{1W}/\partial T = \epsilon\partial\overline{q}_{1W}/\partial\tau$ are $O(\epsilon^2)$ in (2.22). Substituting (3.1) in (3.3), we obtain slow equations for the two geostrophic modes $\boldsymbol{k}^{G1}$ and $\boldsymbol{k}^{G2}$ as

$$\begin{aligned}&\frac{\mathrm{d}q^{G1}}{\mathrm{d}\tau} + [\boldsymbol{k}^{G1}\boldsymbol{\cdot}(\boldsymbol{k}^{W1}+\boldsymbol{k}^{W2})^{\perp}](\boldsymbol{v}^{W1^*}\boldsymbol{\cdot}\boldsymbol{v}^{W2^*})q^{G2} \\ &\quad -\mathrm{i}(\boldsymbol{k}^{G1}\boldsymbol{\cdot}\boldsymbol{v}^{G2})(\boldsymbol{k}^{W1}+\boldsymbol{k}^{W2})\cdot[(\boldsymbol{k}^{W2}\boldsymbol{\cdot}\boldsymbol{v}^{W1^*})\boldsymbol{v}^{W2^*} + (\boldsymbol{k}^{W1}\boldsymbol{\cdot}\boldsymbol{v}^{W2^*})\boldsymbol{v}^{W1^*}] \\ &\quad -\mathrm{i}\frac{q^{G2}}{\omega}\{(\boldsymbol{k}^{G2}\boldsymbol{\cdot}\boldsymbol{v}^{W1^*})(\boldsymbol{k}^{G1}\boldsymbol{\cdot}\boldsymbol{v}^{W2^*}) - (\boldsymbol{k}^{G2}\boldsymbol{\cdot}\boldsymbol{v}^{W2^*})(\boldsymbol{k}^{G1}\boldsymbol{\cdot}\boldsymbol{v}^{W1^*})\} = 0,\end{aligned} \tag{3.4a}$$

$$\frac{\mathrm{d}q^{G2}}{\mathrm{d}\tau} - [\boldsymbol{k}^{G2}\cdot(\boldsymbol{k}^{W1}+\boldsymbol{k}^{W2})^{\perp}](\boldsymbol{v}^{W1}\cdot\boldsymbol{v}^{W2})q^{G1} \\ - \mathrm{i}(\boldsymbol{k}^{G2}\cdot\boldsymbol{v}^{G1})(\boldsymbol{k}^{W1}+\boldsymbol{k}^{W2})\cdot[(\boldsymbol{k}^{W2}\cdot\boldsymbol{v}^{W1})\boldsymbol{v}^{W2}+(\boldsymbol{k}^{W1}\cdot\boldsymbol{v}^{W2})\boldsymbol{v}^{W1}] \\ - \mathrm{i}\frac{q^{G1}}{\omega}\{(\boldsymbol{k}^{G2}\cdot\boldsymbol{v}^{W1})(\boldsymbol{k}^{G1}\cdot\boldsymbol{v}^{W2})-(\boldsymbol{k}^{G2}\cdot\boldsymbol{v}^{W2})(\boldsymbol{k}^{G1}\cdot\boldsymbol{v}^{W1})\}=0. \tag{3.4b}$$

The energy equation for the two geostrophic modes considered is

$$\frac{\mathrm{d}E_G}{\mathrm{d}\tau} = \frac{\mathrm{d}}{\mathrm{d}\tau}[(1+|\boldsymbol{k}^{G1}|^2)|h^{G1}|^2+(1+|\boldsymbol{k}^{G2}|^2)|h^{G2}|^2] \\ = \mathrm{i}\{|\boldsymbol{k}^{G2}|^2-|\boldsymbol{k}^{G1}|^2\}\left\{\mathrm{i}[\boldsymbol{k}^{G1}\cdot(\boldsymbol{k}^{W1}+\boldsymbol{k}^{W2})^{\perp}](\boldsymbol{v}^{W1}\cdot\boldsymbol{v}^{W2}) \right. \\ \left. +\frac{1}{\omega}\{(\boldsymbol{k}^{G2}\cdot\boldsymbol{v}^{W2})(\boldsymbol{k}^{G1}\cdot\boldsymbol{v}^{W1})-(\boldsymbol{k}^{G2}\cdot\boldsymbol{v}^{W1})(\boldsymbol{k}^{G1}\cdot\boldsymbol{v}^{W2})\}\right\} h^{G1}h^{G2^*}+\text{c.c.} \tag{3.5}$$

Hence, all geostrophic modes that satisfy $|\boldsymbol{k}^{G1}|^2 \neq |\boldsymbol{k}^{G2}|^2$ are potentially capable of exchanging energy with the waves via four-wave resonance (see figure 1). For example, consider the special case $\boldsymbol{k}^{W1}=\boldsymbol{k}^{W2}=\boldsymbol{k}$, $\boldsymbol{k}^{G1}=\boldsymbol{k}^{\perp}$ and $\boldsymbol{k}^{G2}=\boldsymbol{k}^{\perp}+2\boldsymbol{k}$, where $\boldsymbol{k}^{\perp}=\hat{\boldsymbol{z}}\times\boldsymbol{k}$. For this system, (3.5) becomes

$$\frac{\mathrm{d}E_G}{\mathrm{d}\tau} = 8k^4 h^{W1}h^{W2}h^{G1}h^{G2^*}+\text{c.c.} \tag{3.6}$$

However, such an example is highly artificial and difficult to realize in practice, since an arbitrary set of two geostrophic modes will excite new geostrophic modes by triad interactions and transfer energy to those modes for time scales $T \sim O(1)$. As a result, a simple set of four amplitude equations that demonstrate energetic interaction between waves and geostrophic modes does not seem plausible.

3.2. *A special set of geostrophic modes*

A special set of geostrophic modes that deserve attention are modes that satisfy $|\boldsymbol{k}^{G1}|^2 = |\boldsymbol{k}^{G2}|^2$. These geostrophic modes are special since in the absence of other modes they do not excite new geostrophic modes by triad interaction due to the fact that $\boldsymbol{v}_G\cdot\nabla q_0=0$ for this set of modes (see appendix D for the proof). Thus, if we consider a system consisting of two such geostrophic modes and two wave modes of the same frequency (such that triad wave–geostrophic–wave interactions are absent), the leading nonlinear interactions would be quadruple interactions. The wave modes are expected to catalyse energy exchange between geostrophic modes in such a system. Of course, as follows from (3.5), these geostrophic modes will preserve their total energy.

We derive an example to demonstrate this mechanism by considering two waves (W1 and W2) and two geostrophic modes (G1 and G2). The waves are chosen from opposite branches but have the same wavenumber, $\boldsymbol{k}^{W1}=\boldsymbol{k}^{W2}=\boldsymbol{k}$. The geostrophic modes have wavenumbers $\boldsymbol{k}^{G1}=2\boldsymbol{k}^{\perp}+\boldsymbol{k}$ and $\boldsymbol{k}^{G2}=2\boldsymbol{k}^{\perp}-\boldsymbol{k}$. Such a set of wavenumbers with $\boldsymbol{k}$ as the only free parameter is a conscious choice to simplify the algebraic calculations involved in constructing the reduced system. For this special choice of two waves and two geostrophic modes there are no triad interactions. Since four-wave resonant interactions are expected to appear due to cubic nonlinearity at $O(\epsilon^2)$ in the asymptotic expansions, the slow time is chosen to be $\tau=\epsilon^2 t$. The slowly evolving amplitude equations for these four modes are given by (details of the derivation are given in appendix E)

$$\frac{\mathrm{d}h^{W1}}{\mathrm{d}\tau}+\mathrm{i}c_{ggw}h^{G1}h^{G2^*}h^{W2^*}+\mathrm{i}(c_{ww}|h^{W1}|^2+c_{g1}|h^{G1}|^2+c_{g2}|h^{G2}|^2)h^{W1}=0, \quad (3.7a)$$

$$\frac{\mathrm{d}h^{W2}}{\mathrm{d}\tau}-\mathrm{i}c_{ggw}h^{G1}h^{G2^*}h^{W1^*}-\mathrm{i}(c_{ww}|h^{W2}|^2+c_{g1}|h^{G1}|^2+c_{g2}|h^{G2}|^2)h^{W2}=0, \quad (3.7b)$$

$$\frac{\mathrm{d}h^{G1}}{\mathrm{d}\tau}+c_{wwg}h^{G2}h^{W1}h^{W2}+c_{gg}|h^{G2}|^2h^{G1}=0, \quad (3.7c)$$

$$\frac{\mathrm{d}h^{G2}}{\mathrm{d}\tau}-c_{wwg}h^{G1}h^{W1^*}h^{W2^*}-c_{gg}|h^{G1}|^2h^{G2}=0. \quad (3.7d)$$

The constants above are given by

$$\left.\begin{aligned} c_{ggw}=4k^2(9k^4+2k^2-3)/3\omega^3, \quad c_{ww}=3\omega k^2,\\ c_{g1}=c_{g2}=(312k^4+128k^2-152)k^2/42\omega^3,\\ c_{wwg}=4(k^2-3)k^2/(1+5k^2), \quad c_{gg}=-24k^4. \end{aligned}\right\} \quad (3.8)$$

We then have the following energy equations from (3.7):

$$\begin{aligned} \frac{\mathrm{d}E^{W1}}{\mathrm{d}\tau} &= \frac{2\mathrm{i}\omega^2c_{ggw}}{k^2}h^{G1^*}h^{G2}h^{W1}h^{W2}+\text{c.c.}=-\frac{\mathrm{d}E^{W2}}{\mathrm{d}\tau}\\ &\Rightarrow E_W=E^{W1}+E^{W2}=\frac{2\omega^2}{k^2}|h^{W1}|^2+\frac{2\omega^2}{k^2}|h^{W2}|^2=E_W(0), \end{aligned} \quad (3.9)$$

$$\begin{aligned} \frac{\mathrm{d}E^{G1}}{\mathrm{d}\tau} &= -(1+5k^2)(c_{wwg}h^{G1^*}h^{G2}h^{W1}h^{W2}+c_{gg}|h^{G1}|^2|h^{G2}|^2+\text{c.c.})=-\frac{\mathrm{d}E^{G2}}{\mathrm{d}\tau}\\ &\Rightarrow E_G=E^{G1}+E^{G2}=(1+5k^2)|h^{G1}|^2+(1+5k^2)|h^{G2}|^2=E_G(0). \end{aligned} \quad (3.10)$$

System (3.7) also conserves potential enstrophy,

$$Q(\tau)=Q^{G1}+Q^{G2}=(1+5k^2)^2|h^{G1}|^2+(1+5k^2)^2|h^{G2}|^2=Q(0). \quad (3.11)$$

The main point we wish to make with the help of (3.7) is that the energetic interaction between balanced modes is affected by the presence of waves, although (3.7) conserves the total wave energy and total geostrophic energy separately. For an illustration of this, (3.7) is solved numerically using RK-4 for the initial conditions $h^{W1}=1$, $h^{W2}=2$, $h^{G1}=3$, $h^{G2}=4$ and $|\boldsymbol{k}|=1/4$. Figure 2 shows the energy exchange between the geostrophic modes in the absence of waves, i.e. obtained by solving (3.7*c*,*d*) after setting $c_{wwg}=0$. It is seen that the G1 mode extracts all of the energy from G2, asymptotically gaining all of the energy in the system, while G2 asymptotes towards zero energy, leading to its demise. Figure 3 shows the effect of having waves in the system obtained by solving the complete system (3.7). In this case, the waves continuously transfer energy between the geostrophic modes in such a way that the G2 mode does not lose all of its energy to the G1 mode. As a minor point, we also observe that the geostrophic modes catalyse energy exchanges between the wave modes, as is shown in figure 4. (In the absence of the geostrophic modes, the waves do not interact, as may be inferred from (3.7*a*) and (3.7*b*).)

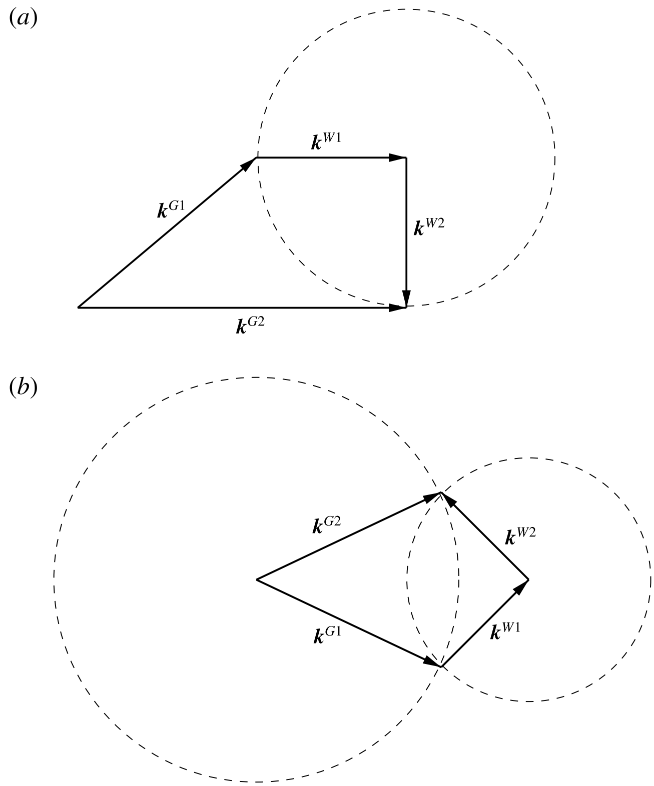

FIGURE 1. (*a*) The general case of quadruple interactions between waves and geostrophic modes in the absence of resonant triads. Two waves (W1 and W2) of the same frequency ($|\boldsymbol{k}^{W1}|^2 = |\boldsymbol{k}^{W2}|^2$) can interact with two geostrophic modes (G1 and G2). The interaction can result in energy exchange between the waves and the geostrophic modes when $|\boldsymbol{k}^{G1}|^2 \neq |\boldsymbol{k}^{G2}|^2$. (*b*) A special case when $|\boldsymbol{k}^{G1}|^2 = |\boldsymbol{k}^{G2}|^2$. Geostrophic modes that fall into this category do not interact among themselves via triad interactions. In such a setting, the waves can catalyse energy exchange between the geostrophic modes, and the geostrophic modes can catalyse energy exchange between the waves. More specifically, the energetic interaction between geostrophic modes is modified by the presence of waves. However, there is no net transfer of energy between geostrophic modes and waves, i.e. the total geostrophic and total wave energies remain separately conserved.

4. Summary

This paper was aimed at investigating the evolution of unbalanced initial data in RSW in the regime $Ro \sim Fr \ll 1$, with the main focus on interactions between slow balanced flow and fast waves. Using the method of multiple time scale asymptotic analysis, we derived an evolution equation for the slow dynamics of RSW valid for $t \lesssim 1/Ro^2$, the well known classical QG equation being valid for $t \lesssim 1/Ro$. Inertio-gravity waves were seen to influence the balanced flow, while spatially homogeneous inertial oscillations did not affect the geostrophic field up to $O(Ro^2)$. In the case of an unbounded domain with compact support for initial data, the wave-balanced flow interaction terms were seen to identically vanish due to fast decay of the wave amplitudes via dispersive propagation, in complete agreement with the results of RZB. In such a set-up, the process of geostrophic adjustment

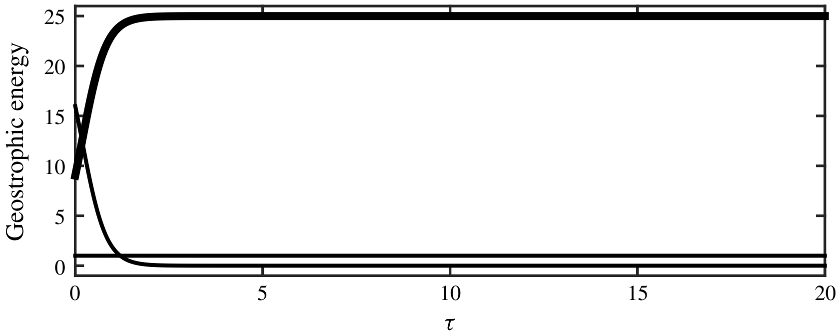

FIGURE 2. Energy of the geostrophic modes (scaled by $1 + 5k^2$) in the absence of wave modes. Thick and thin lines correspond to $|h^{G1}|^2$ and $|h^{G2}|^2$ respectively. The total geostrophic energy (scaled by the initial value) is conserved, as indicated by the straight line.

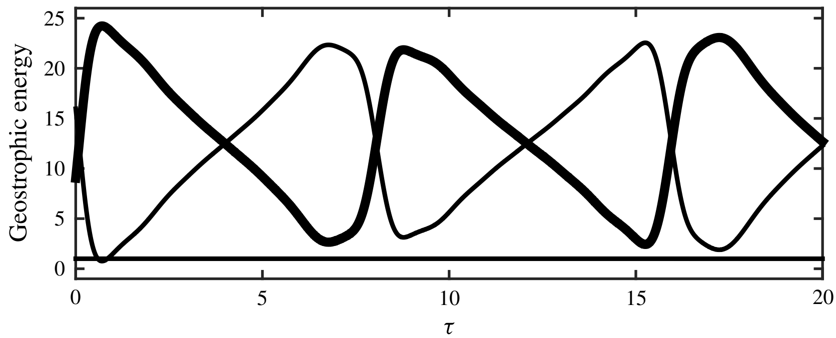

FIGURE 3. Energy of the geostrophic modes (scaled by $1 + 5k^2$) in the presence of wave modes. Thick and thin lines correspond to $|h^{G1}|^2$ and $|h^{G2}|^2$ respectively. The total geostrophic energy (scaled by the initial value) is conserved, as indicated by the straight line.

takes place and the slow balanced flow evolves unaffected by the fast modes, i.e quasi-geostrophy still holds, although the new governing equation is a higher-order balance equation (or an improved QG PV equation) with additional cubic nonlinear geostrophic interaction terms. However, in the case of a general domain that prevents dispersive propagation of waves and in the absence of dissipation (this may be real physical dissipation or artificial/numerical dissipation), the waves can influence the balanced flow. Furthermore, the energy equation corresponding to the slow evolution equation is not closed, implying that the slow dynamics can interact energetically with the fast modes. Thus, in the long-time limit, a splitting between fast and slow energy does not hold. Ertel potential enstrophy was shown to be an exact invariant of the higher-order slow evolution equation, although it contains contributions from the balanced part and the fast dynamics. Thus, fast–slow enstrophy exchanges are also possible in the long-time limit.

Hence, in the long-time limit, $t \gg 1/Ro$, it is clear that an adjustment process leading to the slow evolution of the balanced flow uninfluenced by the fast fields is not possible and the QG framework of describing the evolution of balanced flow without information on wave dynamics breaks down. As a result, if the initial data

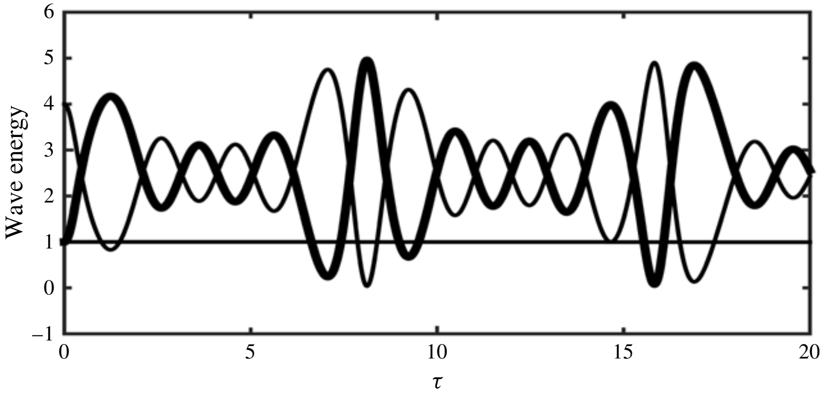

FIGURE 4. Energy of the wave modes (scaled by $2\omega^2/k^2$) in the presence of geostrophic modes. Thick and thin lines correspond to $|h^{W1}|^2$ and $|h^{W2}|^2$ respectively. The total wave energy (scaled by the initial value) is conserved, as indicated by the straight line.

are unbalanced, a slow manifold cannot exist in the long-time limit. Therefore, the central result of this work is that the slow dynamics obtained by filtering fast waves can deviate significantly from the slow dynamics obtained by tracking both fast and slow modes of the full parent model, which in this case is RSW. It is possible that the presence of wave activity influencing the geostrophic cascade as observed by Farge & Sadourny (1989) could be an indication of the early stages of interaction between the waves and the balanced flow. Significant interactions might occur if the dynamics were investigated for longer times with near-inviscid conditions, i.e. in a setting such that numerical dissipation does not considerably affect the dynamics for time scales of the order $t \sim 1/Ro^2$.

The interaction was further investigated in a periodic domain with the goal of constructing a reduced system of slowly evolving amplitude equations. In the long-time limit, $t \sim 1/Ro^2$, significant four-wave resonant interactions are expected to take place. In a system consisting of four modes (chosen such that there are no resonant triads in the system), two waves of the same frequency and two arbitrary geostrophic modes, the total geostrophic energy is not conserved but is exchanged with the fast waves. However, two arbitrary geostrophic modes in general do not form a closed system and can resonantly excite new geostrophic modes by triad interaction. A special case of non-resonant geostrophic modes, i.e. modes that do not excite new geostrophic modes via triad interactions, was chosen to derive the reduced system. Although the total geostrophic energy is conserved in such a system, the presence of waves modifies the energy exchange between the geostrophic modes. That is, the energetic interaction between balanced modes can be very greatly modified in the presence of waves, as was illustrated with a particular example. The geostrophic modes also act as catalysts for energetic interaction between the waves, although the total wave energy is also separately conserved in this reduced system. However, this reduced system only serves as a special example to illustrate that significant fast–slow interactions are possible in the long-time limit. High-accuracy numerical solutions of RSW will have to be undertaken to investigate the long-time state of unbalanced initial data.

Acknowledgements

The author acknowledges the Geophysical Fluid Dynamics – 2014 summer programme at Woods Hole Oceanographic Institution where some of the preliminary analysis that led to this paper was carried out. S. Smith is thanked for interesting discussions on the wave-balanced flow interaction problem, with a focus on periodic domains to investigate the long-time dynamics, for encouraging the author to write this paper and also for useful comments on an earlier version of this paper. The author is indebted to I. Grooms for pointing out some inaccuracies in the first draft and for helping to revise some of the early versions of this paper with critical but constructive comments and suggestions. O. Bühler is thanked for comments and suggestions that helped in the revision of an earlier draft. Three anonymous reviewers are also gratefully acknowledged for comments and suggestions that led to significant improvement of the paper.

Appendix A. Equivalence of two methods of multi-time-scale asymptotic analysis

Here, we show that the asymptotic technique that we used to derive the higher-order slow evolution equation of RSW using a single slow time scale, following Ablowitz (2011), is equivalent to a method that uses two slow time scales. Consider a weakly nonlinear system of the form

$$\frac{\partial u}{\partial t} + \mathscr{L}u + \epsilon\mathscr{B}(u, u) + \epsilon^2\mathscr{N}(u) = 0, \tag{A 1}$$

where $u \in \mathbb{R}$, $\mathscr{L}$ is a linear self-adjoint operator, $\mathscr{B}(u, u)$ is a bilinear operator and $\mathscr{N}(u)$ is an arbitrary nonlinear operator acting on u – these operators may be algebraic or differential (although here we restrict ourselves to the case of u being a scalar field, all of these ideas may be easily extended to higher dimensions).

Two slow time scales

We introduce two slow times $T = \epsilon t$ and $\tau = \epsilon^2 t$ to rewrite (A 1) as

$$\frac{\partial u}{\partial t} + \epsilon\frac{\partial u}{\partial T} + \epsilon^2\frac{\partial u}{\partial \tau} + \mathscr{L}u + \epsilon\mathscr{B}(u, u) + \epsilon^2\mathscr{N}(u) = 0. \tag{A 2}$$

Using an asymptotic expansion of the form $u = u_0 + \epsilon u_1 + \epsilon^2 u_2 + \cdots$, we write the leading-order solution of (A 2) as $u_0 = \mathrm{e}^{-\mathscr{L}t}U(T, \tau)$, where $\mathrm{e}^{-\mathscr{L}t}$ is the linear solution in abstract form, with $\mathscr{L}$ being an algebraic or differential operator and $U(T, \tau)$ being the slowly evolving amplitude. At $O(\epsilon)$ we obtain

$$\frac{\partial u_1}{\partial t} + \mathscr{L}u_1 = -\left\{\frac{\partial u_0}{\partial T} + \mathscr{B}(u_0, u_0)\right\}. \tag{A 3}$$

We use the solvability condition to suppress secular growth of u_1 and obtain the slow evolution equation valid for $T \sim 1$ $(t \sim 1/\epsilon)$,

$$\frac{\partial U}{\partial T} + \mathscr{R}\{\mathscr{B}(u_0, u_0)\} = 0, \tag{A 4}$$

where $\mathscr{R}(\psi)$ is the projection of ψ on the linear solution $\mathrm{e}^{-\mathscr{L}t}$, i.e.

$$\mathscr{R}\{\psi(\boldsymbol{x}, t, T)\} = \lim_{\tilde{T}\to\infty}\frac{1}{\tilde{T}}\int_0^{\tilde{T}} \mathrm{e}^{\mathscr{L}t}\psi(\boldsymbol{x}, t, T)\,\mathrm{d}t. \tag{A 5}$$

By deriving (A 4), we have filtered the fast-time ($t \sim 1$) dynamics from the leading-order solution, a process that is accompanied by several advantages. For example, numerical solution of (A 4) can use larger time steps than those allowed by (A 1). After removing the resonant terms, we can solve (A 3) to obtain u_1. At $O(\epsilon^2)$ of (A 2), we obtain

$$\frac{\partial u_2}{\partial t} + \mathscr{L} u_2 = -\left\{ \frac{\partial u_1}{\partial T} + \frac{\partial u_0}{\partial \tau} + \mathscr{B}(u_0, u_1) + \mathscr{B}(u_1, u_0) + \mathscr{N}(u_0) \right\}, \tag{A 6}$$

where $\mathscr{B}(u_0, u_1) + \mathscr{B}(u_1, u_0)$ represent the terms that arise due to interactions between u_0 and u_1 in abstract form. We use the solvability condition again to eliminate t dependence and obtain the second slow equation valid up to $\tau \sim 1$ ($t \sim 1/\epsilon^2$),

$$\frac{\partial U}{\partial \tau} + \mathscr{R} \left\{ \frac{\partial u_1}{\partial T} + \mathscr{B}(u_0, u_1) + \mathscr{B}(u_1, u_0) + \mathscr{N}(u_0) \right\} = 0. \tag{A 7}$$

Equation (A 7) has variables that evolve on two time scales – T and τ. The T dependence appears above since (1) u_1 obtained by solving (A 3) (after removing resonant terms) depends on u_0; (2) nonlinear interactions, $\mathscr{B}(u_0, u_1)$, $\mathscr{B}(u_1, u_0)$ and $N(u_0)$, have T dependence, u_0 depending on T as given by (A 4). Equation (A 7) is thus coupled to (A 4). In general, writing an evolution equation for the leading-order field valid for $\tau \sim 1$, with the first slow time dependence removed, is a formidable task. Most of the difficulty arises because (A 4) is in general nonlinear and hence difficult to treat exactly. In short, constructing a solvability condition based on (A 4) and using it in (A 7) to filter off T dependence, along the same lines that we adopted in passing from (A 3) to (A 4), is often difficult, except for some special solvable forms of (A 4). In hindsight this is hardly surprising, since we were able to introduce the solvability condition in (A 3) simply because the leading-order equation was linear, due to which we wrote down an exact solution to it. Once we encounter a nonlinear equation at higher order in asymptotics, such as (A 4), except for special cases, exact treatment is not easy.

A pragmatic strategy that bypasses all of these difficulties is to combine the two slow equations and write a single equation that holds for $T \sim 1$ to $T \sim 1/\epsilon$. This is done by summing (A 4) + ϵ (A 7) and defining $\partial/\partial T + \epsilon \partial/\partial \tau \rightarrow \partial/\partial T$, which gives a slow equation

$$\frac{\partial U}{\partial T} + \mathscr{R}\{\mathscr{B}(u_0, u_0)\} + \epsilon \mathscr{R} \left\{ \frac{\partial u_1}{\partial T} + \mathscr{B}(u_0, u_1) + \mathscr{B}(u_1, u_0) + \mathscr{N}(u_0) \right\} = 0, \tag{A 8}$$

which is valid for $T \sim 1/\epsilon$ or $t \sim 1/\epsilon^2$. At this point, one must realize that we have given up the idea of distinguishing between two slow time scales, T and τ. Instead, a single slow time T is allowed to vary from $O(1)$ to $O(1/\epsilon)$ in (A 8). However, it must be noted that in spite of (A 8) containing ϵ, one is at a higher advantage solving (A 8) than solving (A 1) (up to $t \sim 1/\epsilon^2$), which has fast dynamics $t \sim 1$ and hence requires very small time steps. Although the impact of the higher-order terms, i.e. terms multiplied by ϵ, is weak for $T \sim O(1)$ time scales, these terms may change the leading-order field by a significant amount in longer time scales, $T \sim 1/\epsilon$.

A single slow time scale

We now show that one may obtain (A 8) by using a single slow time, i.e. without explicitly introducing two slow time scales, but by modifying the analysis slightly. We introduce a single slow time $T = \epsilon t$ and rewrite (A 2) as

$$\frac{\partial u}{\partial t} + \epsilon \frac{\partial u}{\partial T} + \mathscr{L} u + \epsilon \mathscr{B}(u, u) + \epsilon^2 \mathscr{N}(u) = 0. \tag{A 9}$$

At $O(\epsilon)$ we rewrite (A 4) as

$$\frac{\partial U}{\partial T} + \mathscr{R}\{\mathscr{B}(u_0, u_0)\} = \epsilon \Phi(\boldsymbol{x}, T) + O(\epsilon^2), \tag{A 10}$$

where Φ is a higher-order correction term and will be obtained at the next order. We also retain $O(\epsilon^2)$ on the right-hand side of the above equation to indicate the order of error terms involved and hence to imply that (A 10) is valid for $T \lesssim 1/\epsilon$ but not for $T \gg 1/\epsilon$. At $O(\epsilon^2)$ of asymptotics we obtain

$$\frac{\partial u_2}{\partial t} + \mathscr{L} u_2 = -\left\{ \frac{\partial u_1}{\partial T} + \mathrm{e}^{-\mathscr{L} t} \Phi(\boldsymbol{x}, T) + \mathscr{B}(u_0, u_1) + \mathscr{B}(u_1, u_0) + \mathscr{N}(u_0) \right\}. \tag{A 11}$$

Suppressing secular growth, we obtain

$$\Phi(\boldsymbol{x}, T) + \mathscr{R}\left\{ \frac{\partial u_1}{\partial T} + \mathscr{B}(u_0, u_1) + \mathscr{B}(u_1, u_0) + \mathscr{N}(u_0) \right\} = 0. \tag{A 12}$$

We combine (A 10) and (A 12) to obtain

$$\frac{\partial U}{\partial T} + \mathscr{R}\{\mathscr{B}(u_0, u_0)\} + \epsilon \mathscr{R}\left\{ \frac{\partial u_1}{\partial T} + \mathscr{B}(u_0, u_1) + \mathscr{B}(u_1, u_0) + \mathscr{N}(u_0) \right\} + O(\epsilon^2) = 0. \tag{A 13}$$

Equation (A 13) is the slow equation valid for $t \sim 1/\epsilon^2$, since we eliminated secular growth up to $O(\epsilon^2)$. It should be noted that this is the same equation we obtained using two time scale asymptotics (A 8). Alternatively, if we redefine the time derivative in (A 13) as $\partial/\partial T \to \partial/\partial T + \epsilon \partial/\partial \tau$ and write the corresponding equations at $O(1)$ and $O(\epsilon)$, we obtain (A 4) and (A 7). One may use (A 10) to modify the part of (A 13) multiplied by ϵ. The error involved in the process is $O(\epsilon^2)$ and is insignificant for $T \lesssim 1/\epsilon$. This technique of using a single slow time and eliminating resonance up to $O(\epsilon^2)$ to derive a slow equation for RSW valid up to $t \sim 1/\epsilon^2$ was adopted in this paper. The reader may refer to Ablowitz (2011) for particular examples that take advantage of this method.

Appendix B. Higher-order balance equation

Here, we show that the higher-order balance equation we derived in the absence of waves, (2.24), is equivalent to the improved QG PV equation of RZB.

We ignore waves (i.e. set $\boldsymbol{v}_W = h_W = 0$) and use $\hat{\boldsymbol{z}} \times \zeta_G \boldsymbol{v}_G + \nabla(v_G^2/2) = \boldsymbol{v}_G \boldsymbol{\cdot} \nabla \boldsymbol{v}_G$ in (2.12*a*) to obtain an alternate form of (2.13*c*),

$$\overline{\boldsymbol{v}}_{1G} = -\nabla \frac{\partial h_G}{\partial T} + \hat{\boldsymbol{z}} \times \nabla \overline{h}_1 - (\boldsymbol{v}_G \boldsymbol{\cdot} \nabla)(\nabla h_G) \tag{B 1a}$$

$$\Rightarrow \quad \overline{q}_{1G} = \hat{\boldsymbol{z}} \boldsymbol{\cdot} (\nabla \times \overline{\boldsymbol{v}}_{1G}) - \overline{h}_1 = (\Delta - 1)\overline{h}_1 - 2\partial[h_{Gx}, h_{Gy}], \tag{B 1b}$$

where (B 1b) is equivalent to (2.14c) and $\partial[f, g] = f_x g_y - f_y g_x$. Now,

$$\begin{aligned}
\nabla \cdot (q_0 \overline{\boldsymbol{v}}_{1G}) &= -q_0 \frac{\partial h_G}{\partial T} + \overline{\boldsymbol{v}}_{1G} \cdot \nabla q_0 \\
&\quad \text{(using a modified form of (2.13}d\text{), } \nabla \cdot \overline{\boldsymbol{v}}_{1G} = -\partial h_G/\partial T\text{)} \\
&= -\frac{\partial}{\partial T}(q_0 h_G) + h_G \frac{\partial q_0}{\partial T} + (\hat{\boldsymbol{z}} \times \nabla \overline{h}_1) \cdot \nabla q_0 - \nabla q_0 \cdot \frac{\partial}{\partial T}(\nabla h_G) \\
&\quad - \nabla q_0 \cdot \{(\boldsymbol{v}_G \cdot \nabla) \nabla h_G\} \quad \text{(using (B 1}a\text{))} \\
&= -\frac{\partial}{\partial T}(q_0 h_G) - h_G \nabla \cdot (q_0 \boldsymbol{v}_G) + (\hat{\boldsymbol{z}} \times \nabla \overline{h}_1) \cdot \nabla q_0 - \frac{\partial}{\partial T}\{\nabla q_0 \cdot \nabla h_G\} \\
&\quad + \left\{ \nabla h_G \cdot \frac{\partial}{\partial T} \nabla q_0 - \nabla q_0 \cdot \{(\boldsymbol{v}_G \cdot \nabla) \nabla h_G\} \right\} + O(\epsilon) \\
&\quad \text{(using (2.17))} \\
&= -\frac{\partial}{\partial T}(q_0 h_G + \nabla q_0 \cdot \nabla h_G) - \nabla \cdot (h_G q_0 \boldsymbol{v}_G) + (\hat{\boldsymbol{z}} \times \nabla \overline{h}_1) \cdot \nabla q_0 \\
&\quad - [\nabla h_G \cdot \nabla (\boldsymbol{v}_G \cdot \nabla q_0) + \nabla q_0 \cdot \{(\boldsymbol{v}_G \cdot \nabla) \nabla h_G\}] + O(\epsilon) \\
&\quad \text{(using (2.17) again).}
\end{aligned} \tag{B 2}$$

After some manipulations, one can show that

$$\nabla h_G \cdot \nabla (\boldsymbol{v}_G \cdot \nabla q_0) = \nabla h_G \cdot (\boldsymbol{v}_G \cdot \nabla) \nabla q_0 + \left\{ \hat{\boldsymbol{z}} \times \nabla \left(\tfrac{1}{2} (\nabla h_G)^2 \right) \right\} \cdot \nabla q_0. \tag{B 3}$$

Substituting (B 3) in (B 2) gives

$$\begin{aligned}
\nabla \cdot (q_0 \overline{\boldsymbol{v}}_{1G}) &= -\frac{\partial}{\partial T}(q_0 h_G + \nabla q_0 \cdot \nabla h_G) - \nabla \cdot (h_G q_0 \boldsymbol{v}_G) + \partial[\overline{h}_1, q_0] \\
&\quad - (\boldsymbol{v}_G \cdot \nabla)(\nabla q_0 \cdot \nabla h_G) - \partial \left[\frac{1}{2} (\nabla h_G)^2, q_0 \right] + O(\epsilon).
\end{aligned} \tag{B 4}$$

Substituting (B 1b) and (B 4) in (2.24) and rearranging gives

$$\begin{gathered}
\frac{\partial Q}{\partial T} + \partial \left[\left\{ (h_G + \epsilon \overline{h}_1) - \epsilon \frac{1}{2} (\nabla h_G)^2 \right\}, Q \right] + O(\epsilon^2) = 0, \\
\text{where } Q = (\Delta - 1)(h_G + \epsilon \overline{h}_1) - 2\epsilon \partial [h_{Gx}, h_{Gy}] - \epsilon q_0 h_G - \epsilon \nabla q_0 \cdot \nabla h_G.
\end{gathered} \tag{B 5}$$

One may further combine the slow height fields as $\overline{h} = h_G + \epsilon \overline{h}_1$ to rewrite the above equation as

$$\begin{gathered}
\frac{\partial Q}{\partial T} + \partial \left[\left\{ \overline{h} - \epsilon \frac{1}{2} (\nabla \overline{h})^2 \right\}, Q \right] + O(\epsilon^2) = 0, \\
\text{where } Q = (\Delta - 1)\overline{h} - 2\epsilon \partial [\overline{h}_x, \overline{h}_y] - \epsilon \overline{h} (\Delta - 1) \overline{h} - \epsilon \nabla \overline{h} \cdot \nabla \{(\Delta - 1)\overline{h}\} + O(\epsilon^2).
\end{gathered} \tag{B 6}$$

The error involved in the operation is $O(\epsilon^2)$, which is insignificant for time scales $T \lesssim O(1/\epsilon)$. The above equation is the improved QG PV equation (3.72) of RZB, which was also derived by Allen (1993) and Warn *et al.* (1995).

Appendix C. Derivation of the slow energy equation

Here, we describe the steps leading to the slow energy equation, (2.35). We spatially average (2.34) to obtain

$$\left\langle h_G\frac{\partial q_0}{\partial T}+\epsilon\frac{\partial}{\partial T}\{h_G(\overline{q}_{1G}+\overline{q}_{1W})\}\right\rangle \\ -\epsilon\left\langle \overline{q}_{1G}\frac{\partial h_G}{\partial T}+q_0\boldsymbol{\nabla}h_G\boldsymbol{\cdot}\overline{\boldsymbol{v}}_{1G}+\overline{q}_{1W}\frac{\partial h_G}{\partial T}+q_0\boldsymbol{\nabla}h_G\boldsymbol{\cdot}\overline{\boldsymbol{v}}_{1W}+\overline{(\boldsymbol{\nabla}h_G\boldsymbol{\cdot}\boldsymbol{v}_W)q'_{1W}}\right\rangle \\ +O(\epsilon^2)=0. \tag{C 1}$$

We now observe that

$$\begin{aligned} h_G\frac{\partial q_0}{\partial T} &= h_G\frac{\partial}{\partial T}(\Delta h_G-h_G)=-\frac{\partial}{\partial T}\left(\frac{\boldsymbol{v}_G^2}{2}+\frac{h_G^2}{2}\right)+\boldsymbol{\nabla}\boldsymbol{\cdot}\left(h_G\boldsymbol{\nabla}\frac{\partial h_G}{\partial T}\right) \\ &\Rightarrow \left\langle h_G\frac{\partial q_0}{\partial T}\right\rangle=-\frac{\partial}{\partial T}\left\langle\left(\frac{\boldsymbol{v}_G^2}{2}+\frac{h_G^2}{2}\right)\right\rangle. \end{aligned} \tag{C 2}$$

We use (2.14) to obtain

$$\begin{aligned} h_G\overline{q}_{1G} &= \hat{\boldsymbol{z}}\boldsymbol{\cdot}(h_G\boldsymbol{\nabla}\times\overline{\boldsymbol{v}}_{1G})-h_G\overline{h}_1=\hat{\boldsymbol{z}}\boldsymbol{\cdot}\boldsymbol{\nabla}\times(h_G\overline{\boldsymbol{v}}_{1G})-\hat{\boldsymbol{z}}\boldsymbol{\cdot}(\boldsymbol{\nabla}h_G\times\overline{\boldsymbol{v}}_{1G})-h_G\overline{h}_1 \\ &= \hat{\boldsymbol{z}}\boldsymbol{\cdot}\boldsymbol{\nabla}\times(h_G\overline{\boldsymbol{v}}_{1G})-(\boldsymbol{v}_G\boldsymbol{\cdot}\overline{\boldsymbol{v}}_{1G}+h_G\overline{h}_1) \\ &\Rightarrow \langle h_G\overline{q}_{1G}\rangle=-\langle\boldsymbol{v}_G\boldsymbol{\cdot}\overline{\boldsymbol{v}}_{1G}+h_G\overline{h}_1\rangle \end{aligned} \tag{C 3}$$

and

$$\begin{aligned} h_G\overline{q}_{1W} &= \hat{\boldsymbol{z}}\boldsymbol{\cdot}(h_G\boldsymbol{\nabla}\times\overline{\boldsymbol{v}}_{1W})=\hat{\boldsymbol{z}}\boldsymbol{\cdot}\boldsymbol{\nabla}\times(h_G\overline{\boldsymbol{v}}_{1W})-\hat{\boldsymbol{z}}\boldsymbol{\cdot}(\boldsymbol{\nabla}h_G\times\overline{\boldsymbol{v}}_{1W}) \\ &= \hat{\boldsymbol{z}}\boldsymbol{\cdot}\boldsymbol{\nabla}\times(h_G\overline{\boldsymbol{v}}_{1W})-\boldsymbol{v}_G\boldsymbol{\cdot}\overline{\boldsymbol{v}}_{1W} \\ &\Rightarrow \langle h_G\overline{q}_{1W}\rangle=-\langle\boldsymbol{v}_G\boldsymbol{\cdot}\overline{\boldsymbol{v}}_{1W}\rangle. \end{aligned} \tag{C 4}$$

Equation (2.14c) is used for

$$\begin{aligned} \frac{\partial h_G}{\partial T}\overline{q}_{1G} &= \frac{\partial h_G}{\partial T}\Delta\overline{h}_1-\frac{\partial h_G}{\partial T}\overline{h}_1+\frac{\partial h_G}{\partial T}\Delta\left(\frac{\boldsymbol{v}_G^2}{2}\right)-\frac{\partial h_G}{\partial T}\boldsymbol{\nabla}\boldsymbol{\cdot}(\zeta_G\boldsymbol{\nabla}h_G) \\ &= \overline{h}_1\Delta\frac{\partial h_G}{\partial T}-\overline{h}_1\frac{\partial h_G}{\partial T}+\boldsymbol{\nabla}\boldsymbol{\cdot}\left(\frac{\partial h_G}{\partial T}\boldsymbol{\nabla}\overline{h}_1-\overline{h}_1\boldsymbol{\nabla}\frac{\partial h_G}{\partial T}\right)+\left(\frac{\boldsymbol{v}_G^2}{2}\right)\Delta\frac{\partial h_G}{\partial T} \\ &\quad +\boldsymbol{\nabla}\boldsymbol{\cdot}\left(\frac{\partial h_G}{\partial T}\boldsymbol{\nabla}\left(\frac{\boldsymbol{v}_G^2}{2}\right)-\left(\frac{\boldsymbol{v}_G^2}{2}\right)\boldsymbol{\nabla}\frac{\partial h_G}{\partial T}\right) \\ &\quad -\boldsymbol{\nabla}\boldsymbol{\cdot}\left(\zeta_G\frac{\partial h_G}{\partial T}\boldsymbol{\nabla}h_G\right)+\zeta_G\frac{\partial}{\partial T}\left(\frac{\boldsymbol{v}_G^2}{2}\right) \\ &\Rightarrow \left\langle\frac{\partial h_G}{\partial T}\overline{q}_{1G}\right\rangle=\left\langle\overline{h}_1\frac{\partial q_0}{\partial T}+\frac{\partial}{\partial T}\left(\zeta_G\frac{\boldsymbol{v}_G^2}{2}\right)\right\rangle \end{aligned} \tag{C 5}$$

and (2.13c) is used to obtain

$$
\begin{aligned}
q_0\nabla h_G\boldsymbol{\cdot}\overline{\boldsymbol{v}}_{1G} &= -q_0\boldsymbol{v}_G\boldsymbol{\cdot}\left\{\frac{\partial \boldsymbol{v}_G}{\partial T}+\nabla\left(\overline{h}_1+\frac{\boldsymbol{v}_G^2}{2}\right)\right\}\\
&= -q_0\frac{\partial}{\partial T}\left(\frac{\boldsymbol{v}_G^2}{2}\right)-\nabla\boldsymbol{\cdot}\left\{q_0\boldsymbol{v}_G\left(\overline{h}_1+\frac{\boldsymbol{v}_G^2}{2}\right)\right\}+\left(\overline{h}_1+\frac{\boldsymbol{v}_G^2}{2}\right)\nabla\boldsymbol{\cdot}(q_0\boldsymbol{v}_G)\\
&= -\frac{\partial}{\partial T}\left(q_0\frac{\boldsymbol{v}_G^2}{2}\right)-\overline{h}_1\frac{\partial q_0}{\partial T}-\nabla\boldsymbol{\cdot}\left\{q_0\boldsymbol{v}_G\left(\overline{h}_1+\frac{\boldsymbol{v}_G^2}{2}\right)\right\}+O(\epsilon)\\
&\quad \text{(using (2.17))}\\
&\Rightarrow \langle q_0\nabla h_G\boldsymbol{\cdot}\overline{\boldsymbol{v}}_{1G}\rangle = -\left\langle\frac{\partial}{\partial T}\left(q_0\frac{\boldsymbol{v}_G^2}{2}\right)+\overline{h}_1\frac{\partial q_0}{\partial T}\right\rangle+O(\epsilon).
\end{aligned}
\tag{C 6}
$$

We use (2.14b) for

$$
\begin{aligned}
\frac{\partial h_G}{\partial T}\overline{q}_{1W} &= \frac{\partial h_G}{\partial T}\Delta\left(\frac{\overline{\boldsymbol{v}_W^2}}{2}\right)-\frac{\partial h_G}{\partial T}\overline{h_W^2}-\hat{\boldsymbol{z}}\boldsymbol{\cdot}\left(\overline{\nabla h_W\times\boldsymbol{v}_W}\frac{\partial h_G}{\partial T}\right)\\
&= \left(\frac{\overline{\boldsymbol{v}_W^2}}{2}\right)\frac{\partial}{\partial T}\Delta h_G+\nabla\boldsymbol{\cdot}\left(\frac{\partial h_G}{\partial T}\nabla\left(\frac{\overline{\boldsymbol{v}_W^2}}{2}\right)-\left(\frac{\overline{\boldsymbol{v}_W^2}}{2}\right)\nabla\frac{\partial h_G}{\partial T}\right)-\frac{\partial h_G}{\partial T}\overline{h_W^2}\\
&\quad -\left\{\hat{\boldsymbol{z}}\boldsymbol{\cdot}\nabla\times\left(\frac{\partial h_G}{\partial T}\overline{h_W\boldsymbol{v}_W}\right)-\frac{\partial h_G}{\partial T}\overline{h_W^2}-\frac{\partial \boldsymbol{v}_G}{\partial T}\boldsymbol{\cdot}\overline{h_W\boldsymbol{v}_W}\right\}\\
&\Rightarrow \left\langle\frac{\partial h_G}{\partial T}\overline{q}_{1W}\right\rangle = \left\langle\left(\frac{\overline{\boldsymbol{v}_W^2}}{2}\right)\frac{\partial \zeta_G}{\partial T}+\frac{\partial \boldsymbol{v}_G}{\partial T}\cdot\overline{h_W\boldsymbol{v}_W}\right\rangle
\end{aligned}
\tag{C 7}
$$

and (2.13b) to obtain

$$
\begin{aligned}
\nabla h_G\boldsymbol{\cdot}q_0\overline{\boldsymbol{v}}_{1W} &= -q_0\nabla h_G\boldsymbol{\cdot}\overline{h_W\boldsymbol{v}_W}-q_0\boldsymbol{v}_G\boldsymbol{\cdot}\nabla\left(\frac{\overline{\boldsymbol{v}_W^2}}{2}\right)\\
&= -q_0\nabla h_G\boldsymbol{\cdot}\overline{h_W\boldsymbol{v}_W}+\left(\frac{\overline{\boldsymbol{v}_W^2}}{2}\right)\nabla\boldsymbol{\cdot}(q_0\boldsymbol{v}_G)-\nabla\boldsymbol{\cdot}\left(q_0\boldsymbol{v}_G\frac{\overline{\boldsymbol{v}_W^2}}{2}\right)\\
&\Rightarrow \langle\nabla h_G\boldsymbol{\cdot}q_0\overline{\boldsymbol{v}}_{1W}\rangle = -\left\langle q_0\nabla h_G\boldsymbol{\cdot}\overline{h_W\boldsymbol{v}_W}+\frac{\partial q_0}{\partial T}\left(\frac{\overline{\boldsymbol{v}_W^2}}{2}\right)\right\rangle+O(\epsilon),
\end{aligned}
\tag{C 8}
$$

where we used (2.17). Finally, we use (2.16b) to obtain

$$
\overline{(\nabla h_G\boldsymbol{\cdot}\boldsymbol{v}_W)q'_{1W}} = q_0\nabla h_G\boldsymbol{\cdot}\overline{(h_W\boldsymbol{v}_W)}-\overline{(\nabla h_G\boldsymbol{\cdot}\boldsymbol{v}_W)\nabla q_0\boldsymbol{\cdot}\int^t \boldsymbol{v}_W\,\mathrm{d}t}. \tag{C 9}
$$

Substituting (C 2)–(C 9) in (C 1) and simplifying gives the energy equation (2.35).

Appendix D. Non-resonant geostrophic modes in the periodic domain

Here, we prove that geostrophic modes having the same wavenumber vector magnitude do not excite new modes via triad interaction since $\boldsymbol{v}_G\boldsymbol{\cdot}\nabla q=0$ for these modes. We note that for a particular mode,

$$q=(\Delta-1)h_G \Rightarrow q_k=-(1+|\boldsymbol{k}|^2)h_k, \tag{D 1}$$

$$\boldsymbol{v}_G=\hat{\boldsymbol{z}}\times\nabla h_G \Rightarrow \boldsymbol{v}_k=\mathrm{i}h_k\boldsymbol{k}^{\perp}. \tag{D 2}$$

Consider two geostrophic modes, $(\boldsymbol{v}_G, q)=(\boldsymbol{v}_1, q_1)\mathrm{e}^{\boldsymbol{k}_1\cdot\boldsymbol{x}}+(\boldsymbol{v}_2, q_2)\mathrm{e}^{\boldsymbol{k}_2\cdot\boldsymbol{x}}+\text{c.c.}$ Then,

$$\begin{aligned}
&\boldsymbol{v}_G\boldsymbol{\cdot}\nabla q\\
&=\mathrm{i}\{(\boldsymbol{k}_1\boldsymbol{\cdot}\boldsymbol{v}_2)q_1+(\boldsymbol{k}_2\boldsymbol{\cdot}\boldsymbol{v}_1)q_2\}\mathrm{e}^{\mathrm{i}(\boldsymbol{k}_1+\boldsymbol{k}_2)\cdot\boldsymbol{x}}+\mathrm{i}\{(\boldsymbol{k}_1\boldsymbol{\cdot}\boldsymbol{v}_2^*)q_1-(\boldsymbol{k}_2\boldsymbol{\cdot}\boldsymbol{v}_1)q_2^*\}\mathrm{e}^{\mathrm{i}(\boldsymbol{k}_1-\boldsymbol{k}_2)\cdot\boldsymbol{x}}+\text{c.c.}\\
&=-\{(|\boldsymbol{k}_1|^2-|\boldsymbol{k}_2|^2)(\boldsymbol{k}_1\boldsymbol{\cdot}\boldsymbol{k}_2^{\perp})\}\{h_1h_2\mathrm{e}^{\mathrm{i}(\boldsymbol{k}_1+\boldsymbol{k}_2)\cdot\boldsymbol{x}}-h_1h_2^*\mathrm{e}^{\mathrm{i}(\boldsymbol{k}_1-\boldsymbol{k}_2)\cdot\boldsymbol{x}}\}+\text{c.c.},
\end{aligned} \tag{D 3}$$

where we used (D 2). Thus, there will be no triad interactions, and $\boldsymbol{v}_G\boldsymbol{\cdot}\nabla q=0$ if

$$|\boldsymbol{k}_1|^2=|\boldsymbol{k}_2|^2 \quad \text{or } \boldsymbol{k}_1=c\boldsymbol{k}_2, \tag{D 4}$$

where c is an arbitrary constant.

Appendix E. Derivation of the amplitude equations for quadruple wave-balanced mode interactions

We derive amplitude equations for a system consisting of two geostrophic and two wave modes. Since we will not encounter triad interactions for this reduced system, we let the slow time be $\tau=\epsilon^2 t$ and rewrite (2.2) as

$$\frac{\partial\boldsymbol{v}}{\partial t}+\hat{\boldsymbol{z}}\times\boldsymbol{v}+\nabla h+\epsilon\boldsymbol{v}\boldsymbol{\cdot}\nabla\boldsymbol{v}+\epsilon^2\frac{\partial\boldsymbol{v}}{\partial\tau}=0, \tag{E 1a}$$

$$\frac{\partial h}{\partial t}+\nabla\boldsymbol{\cdot}\boldsymbol{v}+\epsilon\nabla\boldsymbol{\cdot}(h\boldsymbol{v})+\epsilon^2\frac{\partial h}{\partial\tau}=0, \tag{E 1b}$$

$$\frac{\partial q}{\partial t}+\epsilon\nabla\boldsymbol{\cdot}(q\boldsymbol{v})+\epsilon^2\frac{\partial q}{\partial\tau}=0. \tag{E 1c}$$

$O(1)$ *equations*

Equations (2.5) form the $O(1)$ equations. We write the solution of this system as

$$\begin{aligned}
(\boldsymbol{v}_0, h_0)\ =\ &(\boldsymbol{v}^{G1}, h^{G1})\exp\{\mathrm{i}(2\boldsymbol{k}^{\perp}+\boldsymbol{k})\boldsymbol{\cdot}\boldsymbol{x}\}+(\boldsymbol{v}^{G2}, h^{G2})\exp\{\mathrm{i}(2\boldsymbol{k}^{\perp}-\boldsymbol{k})\boldsymbol{\cdot}\boldsymbol{x}\}\\
&+(\boldsymbol{v}^{W1}, h^{W1})\exp\{\mathrm{i}(\boldsymbol{k}\boldsymbol{\cdot}\boldsymbol{x}-\omega t)\}+(\boldsymbol{v}^{W2}, h^{W2})\exp\{\mathrm{i}(\boldsymbol{k}\boldsymbol{\cdot}\boldsymbol{x}+\omega t)\}+\text{c.c.}
\end{aligned} \tag{E 2}$$

$O(\epsilon)$ *equations*

The equations at this order are

$$\frac{\partial\boldsymbol{v}_1}{\partial t}+\hat{\boldsymbol{z}}\times\boldsymbol{v}_1+\nabla h_1=-\boldsymbol{v}_0\boldsymbol{\cdot}\nabla\boldsymbol{v}_0, \tag{E 3a}$$

$$\frac{\partial h_1}{\partial t}+\nabla\boldsymbol{\cdot}\boldsymbol{v}_1=-\nabla\boldsymbol{\cdot}(h_0\boldsymbol{v}_0), \tag{E 3b}$$

$$\frac{\partial q_1}{\partial t}=-\nabla\boldsymbol{\cdot}(q_0\boldsymbol{v}_0). \tag{E 3c}$$

These are the same linear equations as (2.5), but now forced by geostrophic–geostrophic, wave–geostrophic and wave–wave interaction terms on the right-hand side. There are no resonant triads and hence the particular solution of (E 3) is obtained after some lengthy but straightforward calculations as

$$\begin{aligned}
\boldsymbol{v}_1 = {} & \{(4\mathrm{i}\omega - 3 - (19 + 12\mathrm{i}\omega)k^2)\boldsymbol{k} + (3\mathrm{i}\omega + 4 + (20 + 23\mathrm{i}\omega)k^2)\boldsymbol{k}^{\perp}\} \\
& \times \frac{h^{G1}h^{W1}}{7\omega k^2} \exp\{\mathrm{i}((2\boldsymbol{k} + 2\boldsymbol{k}^{\perp})\boldsymbol{\cdot}\boldsymbol{x} - \omega t)\} \\
& + \{(4\mathrm{i}\omega + 1 - 15k^2)\boldsymbol{k} + (4 - \mathrm{i}\omega + (4 + 3\mathrm{i}\omega)k^2)\boldsymbol{k}^{\perp}\}\frac{h^{G2}h^{W1}}{3\omega k^2} \exp\{\mathrm{i}(2\boldsymbol{k}^{\perp}\boldsymbol{\cdot}\boldsymbol{x} - \omega t)\} \\
& + \{(1 - 4\mathrm{i}\omega - 15k^2)\boldsymbol{k} + (-(4 + \mathrm{i}\omega) + (3\mathrm{i}\omega - 4)k^2)\boldsymbol{k}^{\perp}\} \\
& \times \frac{h^{G1^*}h^{W1}}{3\omega k^2} \exp\{\mathrm{i}(-2\boldsymbol{k}^{\perp}\boldsymbol{\cdot}\boldsymbol{x} - \omega t)\} \\
& + \{(-(4\mathrm{i}\omega + 3) + (12\mathrm{i}\omega - 19)k^2)\boldsymbol{k} + (3\mathrm{i}\omega - 4 + (23\mathrm{i}\omega - 20)k^2)\boldsymbol{k}^{\perp}\} \\
& \times \frac{h^{G2^*}h^{W1}}{7\omega k^2} \exp\{\mathrm{i}((2\boldsymbol{k} - 2\boldsymbol{k}^{\perp})\boldsymbol{\cdot}\boldsymbol{x} - \omega t)\} \\
& + \{(4\mathrm{i}\omega + 3 - (12\mathrm{i}\omega - 19)k^2)\boldsymbol{k} + (3\mathrm{i}\omega - 4 + (23\mathrm{i}\omega - 20)k^2)\boldsymbol{k}^{\perp}\} \\
& \times \frac{h^{G1}h^{W2}}{7\omega k^2} \exp\{\mathrm{i}((2\boldsymbol{k} + 2\boldsymbol{k}^{\perp})\boldsymbol{\cdot}\boldsymbol{x} + \omega t)\} \\
& + \{(4\mathrm{i}\omega - 1 + 15k^2)\boldsymbol{k} + (-(4 + \mathrm{i}\omega) + (3\mathrm{i}\omega - 4)k^2)\boldsymbol{k}^{\perp}\} \\
& \times \frac{h^{G2}h^{W2}}{3\omega k^2} \exp\{\mathrm{i}(2\boldsymbol{k}^{\perp}\boldsymbol{\cdot}\boldsymbol{x} + \omega t)\} \\
& + \{(-(1 + 4\mathrm{i}\omega) + 15k^2)\boldsymbol{k} + (4 - \mathrm{i}\omega + (3\mathrm{i}\omega + 4)k^2)\boldsymbol{k}^{\perp}\} \\
& \times \frac{h^{G1^*}h^{W2}}{3\omega k^2} \exp\{\mathrm{i}(-2\boldsymbol{k}^{\perp}\boldsymbol{\cdot}\boldsymbol{x} + \omega t)\} \\
& + \{(3 - 4\mathrm{i}\omega + (12\mathrm{i}\omega + 19)k^2)\boldsymbol{k} + (3\mathrm{i}\omega + 4 + (23\mathrm{i}\omega + 20)k^2)\boldsymbol{k}^{\perp}\} \\
& \times \frac{h^{G2^*}h^{W2}}{7\omega k^2} \exp\{\mathrm{i}((2\boldsymbol{k} - 2\boldsymbol{k}^{\perp})\boldsymbol{\cdot}\boldsymbol{x} + \omega t)\} \\
& + 16\mathrm{i}k^2\boldsymbol{k}^{\perp}h^{G1}h^{G2^*} \exp\{\mathrm{i}2\boldsymbol{k}\boldsymbol{\cdot}\boldsymbol{x}\} - 8\mathrm{i}k^2\boldsymbol{k}h^{G1}h^{G2} \exp\{\mathrm{i}4\boldsymbol{k}^{\perp}\boldsymbol{\cdot}\boldsymbol{x}\} \\
& + \{(1 + 2k^2)\omega\boldsymbol{k} - \mathrm{i}\omega^2\boldsymbol{k}^{\perp}\}\frac{h_{W1}^2}{k^2} \exp\{\mathrm{i}(2\boldsymbol{k}\boldsymbol{\cdot}\boldsymbol{x} - 2\omega t)\} \\
& - \{(1 + 2k^2)\omega\boldsymbol{k} + \mathrm{i}\omega^2\boldsymbol{k}^{\perp}\}\frac{h_{W2}^2}{k^2} \exp\{\mathrm{i}(2\boldsymbol{k}\boldsymbol{\cdot}\boldsymbol{x} + 2\omega t)\} \\
& - \frac{2\mathrm{i}\omega^2}{k^2}\boldsymbol{k}^{\perp}h^{W1}h^{W2} \exp\{\mathrm{i}2\boldsymbol{k}\boldsymbol{\cdot}\boldsymbol{x}\} + \text{c.c.} + \frac{2\omega}{k^2}\boldsymbol{k}(|h^{W2}|^2 - |h^{W1}|^2),
\end{aligned} \tag{E 4}$$

$$\begin{aligned}
h_1 = {} & 8(2\omega + \mathrm{i}k^2)\frac{h^{G1}h^{W1}}{7\omega} \exp\{\mathrm{i}((2\boldsymbol{k} + 2\boldsymbol{k}^{\perp})\boldsymbol{\cdot}\boldsymbol{x} - \omega t)\} + 8(\omega - \mathrm{i}) \\
& \times \frac{h^{G2}h^{W1}}{3\omega} \exp\{\mathrm{i}(2\boldsymbol{k}^{\perp}\boldsymbol{\cdot}\boldsymbol{x} - \omega t)\} \\
& + 8(\omega + \mathrm{i})\frac{h^{G1^*}h^{W1}}{3\omega} \exp\{\mathrm{i}(-2\boldsymbol{k}^{\perp}\boldsymbol{\cdot}\boldsymbol{x} - \omega t)\} + 8(2\omega - \mathrm{i}k^2) \\
& \times \frac{h^{G2^*}h^{W1}}{7\omega} \exp\{\mathrm{i}((2\boldsymbol{k} - 2\boldsymbol{k}^{\perp})\boldsymbol{\cdot}\boldsymbol{x} - \omega t)\} \\
& + 8(2\omega - \mathrm{i}k^2)\frac{h^{G1}h^{W2}}{7\omega} \exp\{\mathrm{i}((2\boldsymbol{k} + 2\boldsymbol{k}^{\perp})\boldsymbol{\cdot}\boldsymbol{x} + \omega t)\} + 8(\omega + \mathrm{i})
\end{aligned}$$

$$
\begin{aligned}
&\times \frac{h^{G2}h^{W2}}{3\omega}\exp\{\mathrm{i}(2\boldsymbol{k}^{\perp}\cdot\boldsymbol{x}+\omega t)\} \\
&+8(\omega-\mathrm{i})\frac{h^{G1^*}h^{W2}}{3\omega}\exp\{\mathrm{i}(-2\boldsymbol{k}^{\perp}\cdot\boldsymbol{x}+\omega t)\}+8(2\omega+\mathrm{i}k^2) \\
&\times \frac{h^{G2^*}h^{W2}}{7\omega}\exp\{\mathrm{i}((2\boldsymbol{k}-2\boldsymbol{k}^{\perp})\cdot\boldsymbol{x}+\omega t)\} \\
&+2\omega^2 h_{W1}^2\exp\{\mathrm{i}(2\boldsymbol{k}\cdot\boldsymbol{x}-2\omega t)\}+2\omega^2 h_{W2}^2\exp\{\mathrm{i}(2\boldsymbol{k}\cdot\boldsymbol{x}+2\omega t)\}+\text{c.c.} \qquad (\text{E}\,5)
\end{aligned}
$$

The homogeneous solution of (E 3) can be absorbed in (E 2). Thus, the homogeneous solution up to $O(\epsilon)$ can be written as $(1+\epsilon)\boldsymbol{v}_0$, $(1+\epsilon)h_0$, which is the standard procedure (see, e.g., Ablowitz 2011). We then proceed to $O(\epsilon^2)$, where the above solutions interact with $O(1)$ solutions to induce four-wave resonance.

$O(\epsilon^2)$ equations

Since equations at this order are cumbersome, we write the forcing terms in abstract form as

$$
\begin{aligned}
\frac{\partial \boldsymbol{v}_2}{\partial t}+\hat{\boldsymbol{z}}\times\boldsymbol{v}_2+\nabla h_2 &= F_v^{W1}\exp\{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)\}+F_v^{W2}\exp\{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}+\omega t)\} \\
&\quad +F_v^{G1}\exp\{\mathrm{i}(2\boldsymbol{k}^{\perp}+\boldsymbol{k})\cdot\boldsymbol{x}\}+F_v^{G2}\exp\{\mathrm{i}(2\boldsymbol{k}^{\perp}-\boldsymbol{k})\cdot\boldsymbol{x}\}+\text{NR}, \qquad (\text{E}\,6a)\\
\frac{\partial h_2}{\partial t}+\nabla\cdot\boldsymbol{v}_2 &= F_h^{W1}\exp\{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)\}+F_h^{W2}\exp\{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}+\omega t)\} \\
&\quad +F_h^{G1}\exp\{\mathrm{i}(2\boldsymbol{k}^{\perp}+\boldsymbol{k})\cdot\boldsymbol{x}\}+F_h^{G2}\exp\{\mathrm{i}(2\boldsymbol{k}^{\perp}-\boldsymbol{k})\cdot\boldsymbol{x}\}+\text{NR}, \qquad (\text{E}\,6b)\\
\frac{\partial q_2}{\partial t} &= F_q^{G1}\exp\{\mathrm{i}(2\boldsymbol{k}^{\perp}+\boldsymbol{k})\cdot\boldsymbol{x}\}+F_q^{G2}\exp\{\mathrm{i}(2\boldsymbol{k}^{\perp}-\boldsymbol{k})\cdot\boldsymbol{x}\}+\text{NR}, \qquad (\text{E}\,6c)
\end{aligned}
$$

where NR above refers to non-resonant terms. Setting the resonant forcing terms of (E 6c) to zero gives slowly evolving amplitude equations for the geostrophic modes. Alternatively, substituting (E 2) in (2.22), ignoring resonant triads and further simplifications lead us to the same equations for the geostrophic modes. To obtain amplitude equations for the waves, we rewrite (E 6a) and (E 6b) in the form

$$
\left(\frac{\partial}{\partial t^2}+1-\Delta\right)v_2=\tilde{F}_v^{W1}\exp\{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)\}+\tilde{F}_v^{W2}\exp\{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}+\omega t)\}+\text{NR}, \qquad (\text{E}\,7)
$$

$$
\left(\frac{\partial}{\partial t^2}+1-\Delta\right)h_2=\tilde{F}_h^{W1}\exp\{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}-\omega t)\}+\tilde{F}_h^{W2}\exp\{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{x}+\omega t)\}+\text{NR}. \qquad (\text{E}\,8)
$$

Setting the resonant terms on the right-hand side to zero gives the amplitude equations for the waves. Thus, we obtain the system of four slowly evolving amplitude equations, (3.7).

REFERENCES

ABLOWITZ, M. J. 2011 *Nonlinear Dispersive Waves – Asymptotic Analysis and Solitons*. Cambridge University Press.

ALLEN, J. S. 1993 Iterated geostrophic intermediate models. *J. Phys. Oceanogr.* **23**, 2447–2461.

BABIN, A., MAHALOV, A. & NICOLAENKO, B. 1997 Global splitting and regularity of rotating shallow-water equations. *Eur. J. Mech.* (*B/Fluids*) **16** (1), 725–754.

BENNEY, D. J. 1962 Non-linear gravity wave interactions. *J. Fluid Mech.* **14**, 577–584.
BLUMEN, W. 1972 Geostrophic adjustment. *Rev. Geophys.* **10**, 485–528.
CHARNEY, J. G. 1948 On the scale of atmospheric motions. *Geofys. Publ. Oslo* **17** (2), 1–17.
CRAIK, A. D. D. 1985 *Wave Interactions and Fluid Flows*. Cambridge University Press.
DEWAR, W. K. & KILLWORTH, P. D. 1995 Do fast gravity waves interact with geostrophic motions? *Deep-Sea Res.* I **42** (7), 1063–1081.
FARGE, M. & SADOURNY, R. 1989 Wave-vortex dynamics in rotating shallow water. *J. Fluid Mech.* **206**, 433–462.
GILL, A. E. 1982 *Atmosphere–Ocean Dynamics*. Academic.
LEITH, C. E. 1980 Nonlinear normal mode initialization and quasi-geostrophic theory. *J. Atmos. Sci.* **37**, 958–968.
LORENZ, E. N. 1980 Attractor sets and quasi-geostrophic equilibrium. *J. Atmos. Sci.* **37**, 1685–1699.
LORENZ, E. N. 1986 On the existence of a slow manifold. *J. Atmos. Sci.* **43**, 1547–1557.
LORENZ, E. N. & KRISHNAMURTY, V. 1987 On the non-existence of a slow manifold. *J. Atmos. Sci.* **44**, 2940–2950.
MAJDA, A. J. 2002 *Introduction to Partial Differential Equations and Waves for the Atmosphere and Ocean – Courant Lecture Notes, Bd. 9*. American Mathematical Society.
MAJDA, A. J. & EMBID, P. 1998 Averaging over fast gravity waves for geophysical flows with unbalanced initial data. *Theor. Comput. Fluid Dyn.* **11**, 155–169.
OBUKHOV, A. M. 1949 On the question of geostrophic wind. *Bull. Acad. Sci. USSR Ser. Geophys. Geograph.* **13** (4), 281–306 (in Russian).
PHILLIPS, O. M. 1960 On the dynamics of unsteady gravity waves of finite amplitude. Part 1. The elementary interactions. *J. Fluid Mech.* **9**, 193–217.
REZNIK, G. M. 2014 Geostrophic adjustment with gyroscopic waves: barotropic fluid without the traditional approximation. *J. Fluid Mech.* **743**, 585–605.
REZNIK, G. M. & GRIMSHAW, R. 2002 Nonlinear geostrophic adjustment in the presence of a boundary. *J. Fluid Mech.* **471**, 257–283.
REZNIK, G. M., ZEITLIN, V. & BEN JELLOUL, M. 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.* **2**, 239–263.
VANNESTE, J. 2013 Balance and spontaneous wave generation in geophysical flows. *Annu. Rev. Fluid Mech.* **45** (1), 147–172.
VANNESTE, J. & YAVNEH, I. 2004 Exponentially small inertia-gravity waves and the breakdown of quasigeostrophic balance. *J. Atmos. Sci.* **61**, 211–223.
VAUTARD, R. & LEGRAS, B. 1986 Invariant manifolds, quasi-geostrophy and initialization. *J. Atmos. Sci.* **43**, 565–584.
WARD, M. L. & DEWAR, W. K. 2010 Scattering of gravity waves by potential vorticity in a shallow-water fluid. *J. Fluid Mech.* **663**, 478–506.
WARN, T. 1986 Statistical mechanical equilibria of the shallow water equations. *Tellus* A **38A**, 1–11.
WARN, T., BOKHOVE, O., SHEPHERD, T. G. & VALLIS, G. K. 1995 Rossby-number expansions, slaving principles and balance dynamics. *Q. J. R. Meteorol. Soc.* **121**, 723–739.
WARN, T. & MENARD, R. 1986 Nonlinear balance and gravity-inertial wave saturation in a simple atmospheric model. *Tellus* A **38A**, 285–294.