

Balance model for equatorial long waves

Article

Published Version

Creative Commons: Attribution-Noncommercial-Share Alike 3.0

Chan, I. H. and Shepherd, T. G. ORCID:
<https://orcid.org/0000-0002-6631-9968> (2013) Balance model
for equatorial long waves. *Journal of Fluid Mechanics*, 725. pp.
55-90. ISSN 0022-1120 doi:
<https://doi.org/10.1017/jfm.2013.146> Available at
<https://centaur.reading.ac.uk/33305/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1017/jfm.2013.146>

Publisher: Cambridge University Press

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Balance model for equatorial long waves

Ian H. Chan^{1,†} and Theodore G. Shepherd^{1,2}

¹Department of Physics, University of Toronto, Toronto, ON, Canada M5S 1A7

²Department of Meteorology, University of Reading, Reading RG6 6BB, UK

(Received 23 August 2012; revised 28 November 2012; accepted 7 March 2013;
first published online 14 May 2013)

Geophysical fluid models often support both fast and slow motions. As the dynamics are often dominated by the slow motions, it is desirable to filter out the fast motions by constructing balance models. An example is the quasi-geostrophic (QG) model, which is used widely in meteorology and oceanography for theoretical studies, in addition to practical applications such as model initialization and data assimilation. Although the QG model works quite well in the mid-latitudes, its usefulness diminishes as one approaches the equator. Thus far, attempts to derive similar balance models for the tropics have not been entirely successful as the models generally filter out Kelvin waves, which contribute significantly to tropical low-frequency variability. There is much theoretical interest in the dynamics of planetary-scale Kelvin waves, especially for atmospheric and oceanic data assimilation where observations are generally only of the mass field and thus do not constrain the wind field without some kind of diagnostic balance relation. As a result, estimates of Kelvin wave amplitudes can be poor. Our goal is to find a balance model that includes Kelvin waves for planetary-scale motions. Using asymptotic methods, we derive a balance model for the weakly nonlinear equatorial shallow-water equations. Specifically we adopt the ‘slaving’ method proposed by Warn *et al.* (*Q. J. R. Meteorol. Soc.*, vol. 121, 1995, pp. 723–739), which avoids secular terms in the expansion and thus can in principle be carried out to any order. Different from previous approaches, our expansion is based on a long-wave scaling and the slow dynamics is described using the height field instead of potential vorticity. The leading-order model is equivalent to the truncated long-wave model considered previously (e.g. Heckley & Gill, *Q. J. R. Meteorol. Soc.*, vol. 110, 1984, pp. 203–217), which retains Kelvin waves in addition to equatorial Rossby waves. Our method allows for the derivation of higher-order models which significantly improve the representation of Rossby waves in the isotropic limit. In addition, the ‘slaving’ method is applicable even when the weakly nonlinear assumption is relaxed, and the resulting nonlinear model encompasses the weakly nonlinear model. We also demonstrate that the method can be applied to more realistic stratified models, such as the Boussinesq model.

Key words: atmospheric flows, shallow water flows, stratified flows

† Email address for correspondence: ianchan@atmosph.physics.utoronto.ca

1. Introduction

The quasi-geostrophic (QG) model has been a cornerstone of modern meteorology and oceanography for over half a century. Not only does it underpin most of our theoretical understanding of large-scale dynamics in the atmosphere and ocean, it is also crucial to producing accurate forecasts. The observed large-scale dynamics is typically dominated by slow Rossby waves and vortical motion, but errors in observations can impart an unacceptably large component of fast motions dominated by inertia-gravity (IG) waves in numerical models. The QG model takes advantage of the separation in time scale in the primitive equations, and projects the dynamics onto a lower dimension ‘slow manifold’, thereby filtering out IG waves (e.g. Warn *et al.* 1995). The QG model belongs to a class of models called balance models.

Balance models are derived by taking advantage of a separation in time scale between the fast and slow dynamics in a system. For example, the traditional QG model assumes that the advective time scale is much larger than the inertial time scale:

$$\frac{L/U}{1/f} \gg 1 \iff Ro = \frac{U}{fL} \ll 1. \quad (1.1)$$

Balance models consist of a set of diagnostic balance relations, which can be viewed as a geometrical description of the slow manifold, together with prognostic equations describing the evolution of the system on the slow manifold; these are respectively exemplified by geostrophic balance and the potential vorticity (PV) evolution equation in the QG model. While advances in computers quickly made the PV equation obsolete as a forecasting model for the atmosphere, the diagnostic balance relations in the QG model remained useful for the adjustment of initial data used in forecasting models to avoid spurious high-frequency oscillations (Daley 1993). Today balance relations are at the heart of data assimilation systems (Parrish & Derber 1992; Fisher 2003), as they provide constraints on both observed and non-observed variables, thereby producing analyses that reflect the largely balanced state of the atmosphere.

While geostrophic balance and its higher-order generalization, Bolin–Charney balance (Charney 1955; Bolin 1955), are immensely useful in the mid-latitudes, their applicability diminishes in the tropics: a singularity develops in geostrophic balance, while Bolin–Charney balance distorts the equatorial wave dynamics (Moura 1976; Gent & McWilliams 1983). The result is that balance relations are not used in data assimilation in the tropics (Derber & Bouttier 1999), which means that mass field (i.e. temperature, pressure) measurements are not used to constrain the wind field; this problem is exacerbated by the relatively scarce wind field measurements in the tropics. Without mass–wind coupling the increments in data assimilation do not properly reflect the balanced aspects of large-scale tropical flows. In ocean data assimilation experiments, unbalanced increments can result in degradation instead of improvement as observations are assimilated (Burgers *et al.* 2002), and can lead to an unrealistically deep overturning circulation near the equator (Bell, Martin & Nichols 2004).

For the atmosphere, the effect of a lack of a proper balance model is evident as reanalyses can give vastly different estimates for the tropics. Examples include: (i) the difference in equatorial zonal wind between the NCEP-NCAR and ECMWF ERA-15 reanalyses is of the same magnitude as the interannual variation (Kistler *et al.* 2001); (ii) the estimation of equatorial Kelvin wave activity in the lower stratosphere given by five different reanalyses differ by more than a factor of three (SPARC International Project Office 2010); and (iii) in a comparison between radiosonde data and four

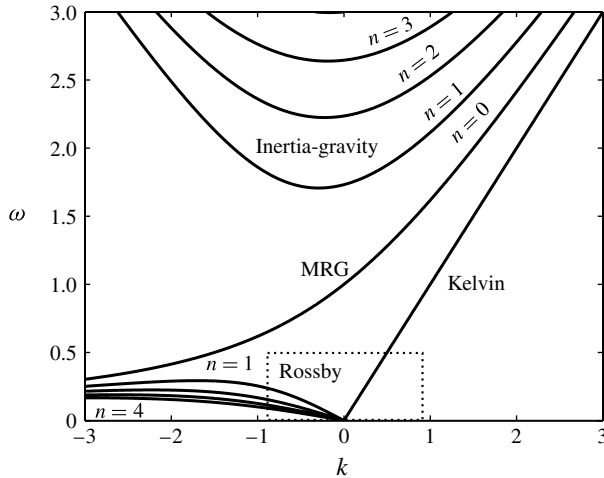


FIGURE 1. Dispersion relation for the linear waves of the equatorial shallow water equation. The scale for the frequency ω is $\sqrt{\beta c}$, while the scale for the zonal wavenumber k is $\sqrt{\beta/c}$. Dotted box indicates the slow dynamics in the long-wave limit.

different reanalyses in the tropical tropopause region, Fujiwara *et al.* (2009) found general agreement in observed zonal windspeed except during the passage of several Kelvin wave packets, suggesting that Kelvin waves may be poorly represented in reanalyses.

We should stress that while Kelvin waves are traditionally often viewed as fast unbalanced motion, they are in fact slow in the long-wave limit. By inspecting the dispersion relations of equatorial waves (cf. figure 1), we can see that in the long-wave regime ($k \ll 1$) there is a clear separation in time scale between Rossby and Kelvin waves and the fast IG waves. As we expect a balance model to accurately capture the slow dynamics (indicated by the dotted box in figure 1), it is essential for a tropical balance model to capture Kelvin waves, in addition to Rossby waves, as they play a central role in tropical low-frequency variability.

While in the tropics Bolin–Charney balance formally holds in the long-wave limit, it does not contain Kelvin waves (Gent & McWilliams 1983). Raymond (1994) pointed out that the inherent difficulty in extending extratropical balanced models to the tropics is that Kelvin waves produce zero perturbation in the PV field, while extratropical balance models are all PV based; Kelvin waves are thus invisible in the balance models. Indeed more recent efforts to derive an equatorial PV-based balance model all suffer from this same issue (Saujani 2005; Theiss & Mohebalhojeh 2009; Verkley & van der Velde 2010). Another hurdle is that unlike the mid-latitudes, where a clear frequency separation exists between Rossby and IG waves, the equatorial region also admits Kelvin and mixed Rossby–gravity (MRG) waves which span the entire frequency range (see figure 1). This leads to ambiguities, as long Kelvin waves are slow while short Kelvin waves are fast, while only westward propagating MRG waves are slow. It is therefore likely that a balance model that admits Kelvin waves will only be valid for long waves.

Indeed the long-wave regime holds promise. The corresponding scaling is equivalent to dropping the acceleration term v_t in the meridional momentum equation for linear equatorial shallow-water equations (SWEs), and the resulting set of equations retains

Kelvin waves (Heckley & Gill 1984). Stevens *et al.* (1990) applied this scaling to arrive at a balance model for SWEs on a sphere for Kelvin and Rossby waves. While these truncations to the SWEs lead to models that are balanced, in the sense that high-frequency IG waves are absent, they prevent the inclusion of higher-order corrections to the balance model, which are important in applications such as model initialization to adequately suppress fast oscillations (Daley 1993).

It should be noted that the long-wave models considered in the literature are characterized by a geostrophic balance between the zonal wind and meridional pressure gradient, and they are thus mathematically similar to the mid-latitude semi-geostrophic model, which is a balance model characterized by a partial geostrophic balance; this similarity supports the view that the long-wave models should be classified as balance models. Indeed, Verkley & van der Velde (2010) argued on this basis that Kelvin waves should be considered as balanced.

Our present study builds on the previous work in the long-wave regime, with the aim of deriving a series of increasingly accurate balance models via asymptotic expansion. Our derivation follows closely the modified asymptotic expansion (‘slaving’ method) proposed by Warn *et al.* (1995). Unlike traditional asymptotic methods where all variables are expanded, the slow variable s used to describe the dynamics on the slow manifold is not expanded, while other fast variables f are then ‘slaved’ onto s via the balance relations:

$$f = U(s; \epsilon) = f_0 + \epsilon f_1 + \epsilon^2 f_2 + \epsilon^3 f_3 + \dots \quad (1.2)$$

The slaving method is thus an asymptotic expansion of the balance relations, which offers two advantages: the method avoids secular growth in the higher-order terms and it allows for the use of variables other than PV to describe the slow dynamics of the system, as long as the variable projects onto the slow modes of the system. Warn *et al.* (1995) successfully applied the method to the f -plane SWEs to recover the QG and Bolin–Charney models based on the Rossby number Ro as the small parameter with PV as the slaving variable. In addition they carried out an alternative expansion using the height perturbation η to describe the slow dynamics, and obtained models that formally have the same accuracy as the QG and Bolin–Charney model.

In this paper we use mass field variables (i.e. η for the SWEs and potential temperature θ for a stratified fluid) as the slaving variable to systematically derive a hierarchy of tropical balance models of increasing accuracy. As Kelvin waves are often characterized by perturbations to the mass field, which is the role η plays in the SWEs, it is reasonable to expect that Kelvin waves will be admitted by the balance model. The added advantage is that the resulting balance equations, which will be in the form $\mathbf{u} = U(\eta)$ or $U(\theta)$, give directly the wind field \mathbf{u} corresponding to temperature observations for a balanced flow.

We first provide a brief review of tropical wave dynamics in §2. The balance models are derived from the equatorial SWEs in the weakly nonlinear limit in §3, with particular attention given to the linear model in order to connect the balance model with existing equatorial wave theory. The accuracy of the balance relations is tested numerically in §4 by considering the accuracy of the wind field inverted from the height field via the linearized model, as well as the use of the weakly nonlinear balance relations to initialize a numerical integration of the SWEs. The weakly nonlinear model is then generalized to a fully nonlinear model in §5. In §6, we shift our focus to the Boussinesq equations to demonstrate that the ‘slaving’ method is applicable to more realistic stratified models.

2. Equatorial wave dynamics

The SWEs are widely used as a model for the atmosphere and oceans, since despite their relative simplicity they retain qualitatively much of the observed features for large-scale atmospheric and oceanic flows. With u and v respectively being the zonal and meridional velocity, and h being the thickness of the fluid, the SWEs are

$$u_t + uu_x + vu_y - fv + gh_x = 0, \quad (2.1a)$$

$$v_t + uv_x + vv_y + fu + gh_y = 0 \quad (2.1b)$$

and

$$h_t + (hu)_x + (hv)_y = 0, \quad (2.1c)$$

where x and y are the zonal and meridional coordinates, f is the Coriolis parameter and g is the gravitational acceleration. The subscripts denote partial derivatives. In the equatorial region the Coriolis parameter can be linearized via the β -plane approximation:

$$f = \beta y, \quad (2.2)$$

with $\beta = 2.3 \times 10^{-11} \text{ m}^{-1} \text{ s}^{-1}$ for the Earth. To non-dimensionalize the above system, first note that for a fluid with mean thickness H , the equatorial Rossby radius of deformation L_R is defined to be $\sqrt{c/\beta}$, where $c = \sqrt{gH}$ is the gravity-wave speed. If we assume that the meridional scale L_y is set by L_R , the scale for the Coriolis term in (2.2) is then $\beta L_y = \sqrt{\beta c}$, and the Rossby and Froude numbers can be defined via:

$$Ro = \frac{U}{fL_x} = \frac{U}{\sqrt{\beta c}L_x} \quad \text{and} \quad Fr = \frac{U}{c}, \quad (2.3)$$

with L_x being the zonal length scale. The ratio of the two numbers gives a horizontal aspect ratio

$$\alpha = \frac{L_y}{L_x} = \frac{Ro}{Fr}. \quad (2.4)$$

We non-dimensionalize the variables via

$$x = L_x \tilde{x}, \quad y = L_y \tilde{y}, \quad t = \frac{1}{\sqrt{\beta c}} \tilde{t}, \quad (2.5a)$$

$$(u, v) = U (\tilde{u}, \tilde{v}) \quad \text{and} \quad h = H(1 + Fr \tilde{\eta}). \quad (2.5b)$$

Dropping the tildes, the non-dimensionalized equations are

$$u_t + Ro \left(uu_x + \frac{1}{\alpha} vu_y \right) - yv + \alpha \eta_x = 0, \quad (2.6a)$$

$$v_t + Ro \left(uv_x + \frac{1}{\alpha} vv_y \right) + yu + \eta_y = 0 \quad (2.6b)$$

and

$$\eta_t + Ro \left((\eta u)_x + \frac{1}{\alpha} (\eta v)_y \right) + \alpha u_x + v_y = 0. \quad (2.6c)$$

2.1. Normal mode solutions to the linear equations

The normal mode solutions to the linearized version of (2.6) were found by Matsuno (1966), and are discussed in various texts such as Gill (1982). A summary of the

results is provided here. Dropping the nonlinear terms in (2.6), and enforcing an isotropic scaling (i.e. taking $\alpha = 1$), we have

$$u_t - yv + \eta_x = 0, \quad (2.7a)$$

$$v_t + yu + \eta_y = 0 \quad (2.7b)$$

and

$$\eta_t + u_x + v_y = 0. \quad (2.7c)$$

We seek normal mode solutions of the form

$$(u, v, \eta) = (\check{u}(y), \check{v}(y), \check{\eta}(y)) \exp(ikx - i\omega t). \quad (2.8)$$

Substituting the above into (2.7), the three equations can be collapsed into a second-order ordinary differential equation in y :

$$\frac{d^2 \check{v}}{dy^2} - \left(\frac{k}{\omega} + k^2 - \omega^2 + y^2 \right) \check{v} = 0, \quad (2.9)$$

which is the parabolic cylinder equation. The standard solutions that remain bounded as $|y| \rightarrow \infty$ are given by Hermite functions:

$$\phi_n = \frac{1}{\sqrt{2^n n!} \sqrt{\pi}} H_n(y) e^{-(1/2)y^2}, \quad (2.10)$$

where H_n is the n th-degree Hermite polynomial. The functions $\{\phi_n\}$ with $n = 0, 1, 2, \dots$ form an orthonormal basis over $(-\infty, \infty)$. For the rest of this paper we use ϕ_n and $H_n e^{-(1/2)y^2}$ interchangeably as they only differ by a constant factor. Here ω and k satisfy a dispersion relation of the form

$$\omega^2 - k^2 - \frac{k}{\omega} = 2n + 1. \quad (2.11)$$

In terms of Hermite functions, \check{v} can be written as

$$\check{v} = i \frac{\omega^2 - k^2}{\omega} H_n e^{-(1/2)y^2}, \quad (2.12a)$$

while

$$\check{\eta} = \left(H'_n - \left(1 + \frac{k}{\omega} \right) y H_n \right) e^{-(1/2)y^2} \quad (2.12b)$$

and

$$\check{u} = \left(\frac{k}{\omega} H'_n - \left(1 + \frac{k}{\omega} \right) y H_n \right) e^{-(1/2)y^2}. \quad (2.12c)$$

The dispersion relation (2.11) is cubic in ω , and there are in general three solutions. The smallest root, ω_R , is identified as the equatorial Rossby wave, and since $\omega_R^2 \ll 1$, the dispersion relation can be approximated by

$$\omega_R = -\frac{k}{2n + 1 + k^2}. \quad (2.13)$$

The two larger roots correspond to equatorial IG waves, with frequency approximately given by

$$\omega_{IG} = \pm \sqrt{2n + 1 + k^2}. \quad (2.14)$$

There are two special wave solutions that need to be considered separately. For $n = 0$, the dispersion relation is satisfied by $\omega = -k$; but since $H'_0 = 0$, equation (2.12) then implies $\check{u} = \check{v} = \check{\eta} = 0$ so this corresponds to a trivial solution (in the literature this is sometimes referred to as the anti-Kelvin wave). The remaining two roots satisfy $\omega_{MRG}^2 - k\omega_{MRG} - 1 = 0$, and these give rise to the MRG waves with

$$\check{u} = \check{\eta} = ye^{-(1/2)y^2} \exp(ikx - i\omega t) \quad \text{and} \quad \check{v} = -i(\omega - k)e^{-(1/2)y^2} \exp(ikx - i\omega t). \quad (2.15)$$

Note that the above expressions are obtained by setting $n = 0$ and $H_0 = 1$ in (2.12), and rescaling the variables. The system (2.7) also admits Kelvin waves given by

$$\check{u} = \check{\eta} = e^{-(1/2)y^2} \exp(ikx - i\omega t) \quad \text{and} \quad \check{v} = 0 \quad \text{with} \quad \omega_K = k. \quad (2.16)$$

2.2. Equatorial long waves

The dispersion relations for the various wave modes are plotted in figure 1. It is evident that in the equatorial region there is a clear time scale separation between Rossby and IG waves, similar to the case in the mid-latitudes; however, the presence of MRG and Kelvin waves in the equatorial region complicates the picture as they span the entire frequency spectrum, and we therefore do not expect motions to be balanced across all zonal scales.

Instead, we may wish to derive a balance model that is valid for certain regimes. A closer look at figure 1 suggests that in the long-wave regime (i.e. $|k| \ll 1$), there is a clear time scale separation between the slower Rossby and Kelvin waves and the faster IG and MRG waves; more precisely, we have

$$\frac{\omega_{R,K}}{\omega_{IG,MRG}} \rightarrow 0 \quad \text{as} \quad k \rightarrow 0. \quad (2.17)$$

The above observation motivates a long-wave scaling, which is achieved formally by defining a small parameter

$$\epsilon = \alpha = L_y/L_x \ll 1. \quad (2.18)$$

When $|k| \sim O(\epsilon) \ll 1$, the Rossby wave frequency can be approximated by

$$\omega_R \approx -\frac{k}{2n+1}, \quad (2.19)$$

while the Kelvin wave frequency, $\omega_K = k$, remains the same. We therefore anticipate the frequency of the slow motion to scale as ϵ in the long-wave regime. To facilitate the derivation of a balance model, we thus rescale time via

$$t = \frac{1}{\epsilon} \tilde{t} \quad (2.20)$$

to put the equations on the slow time scale. With this rescaling and dropping the tilde, we obtain

$$u_t + Fr \left(uu_x + \frac{1}{\epsilon} vu_y \right) - \frac{1}{\epsilon} yv + \eta_x = 0, \quad (2.21a)$$

$$v_t + Fr \left(uv_x + \frac{1}{\epsilon} vv_y \right) + \frac{1}{\epsilon} (yu + \eta_y) = 0 \quad (2.21b)$$

and

$$\eta_t + Fr \left((\eta u)_x + \frac{1}{\epsilon} (\eta v)_y \right) + u_x + \frac{1}{\epsilon} v_y = 0. \quad (2.21c)$$

Equation (2.21b) immediately suggests that to a good approximation, one can ignore v_t and the nonlinear terms in the meridional momentum equation in the long-wave regime, which leaves a simple geostrophic type balance between the zonal wind and meridional pressure gradient.

A long-wave model based on such a truncation has been used previously to study tropical dynamics, and the truncation has been shown to eliminate IG and MRG waves from the SWEs while retaining Rossby and Kelvin waves (see, for example, Heckley & Gill 1984; Stevens *et al.* 1990; Majda 2003; Majda & Klein 2003). While truncation is effective at filtering IG waves, the dynamics of the slow wave modes is inadvertently altered by the truncation, and the truncated terms cannot be systematically reintroduced in the form of higher-order corrections. In §3, we demonstrate how the ‘slaving’ technique, developed by Warn *et al.* (1995), can be used to obtain higher-order generalizations of the long-wave model.

2.3. Conservation laws

We conclude this section with a summary of the conservation laws satisfied by the SWEs (Vallis 2006). Besides the conservation of mass given by (2.21c), the PV (non-dimensionalized by $\sqrt{\beta c}/H$), defined as

$$Q = \frac{Fr(\epsilon v_x - u_y) + y}{1 + Fr \eta}, \quad (2.22)$$

is materially conserved, i.e.

$$\frac{DQ}{Dt} = \left(\frac{\partial}{\partial t} + Fr \left(u \frac{\partial}{\partial x} + \frac{v}{\epsilon} \frac{\partial}{\partial y} \right) \right) Q = 0. \quad (2.23)$$

Another conserved quantity is energy, with the kinetic energy defined as $\mathcal{K} = h(u^2 + v^2)/2$ and the potential energy given by $\mathcal{P} = gh^2/2$. The dimensional equation for total energy $\mathcal{E} = \mathcal{K} + \mathcal{P}$ is

$$\frac{\partial \mathcal{E}}{\partial t} + \nabla \cdot \left(\mathcal{E} \mathbf{u} + \frac{1}{2} gh^2 \mathbf{u} \right) = 0, \quad (2.24)$$

with $\nabla = (\partial_x, \partial_y)$ and $\mathbf{u} = (u, v)$. Upon non-dimensionalization with (2.5) and $\mathcal{E} = gH^2 \tilde{\mathcal{E}}$, the non-dimensional energy is given by

$$\tilde{\mathcal{E}} = Fr^2(1 + Fr \eta) \frac{u^2 + v^2}{2} + \frac{1}{2} (1 + Fr \eta)^2, \quad (2.25)$$

while (2.24) becomes (after dropping the tilde)

$$\frac{\partial \mathcal{E}}{\partial t} + Fr \left(\frac{\partial}{\partial x}, \frac{1}{\epsilon} \frac{\partial}{\partial y} \right) \cdot \left(\mathbf{u} \left(\mathcal{E} + \frac{1}{2} (1 + Fr \eta)^2 \right) \right). \quad (2.26)$$

Finally, the equatorial SWEs also conserve the absolute momentum $\mathcal{M} = h(u - \beta y^2)/2$ (Ripa 1983):

$$\frac{\partial \mathcal{M}}{\partial t} + \nabla \cdot (\mathbf{u} \mathcal{M}) + \frac{\partial}{\partial x} \left(\frac{1}{2} gh^2 \right) = 0. \quad (2.27)$$

Defining $\tilde{\mathcal{M}} = \mathcal{M}/cH$, we have

$$\tilde{\mathcal{M}} = (1 + Fr \eta) \left(Fr u - \frac{1}{2} y^2 \right) \quad (2.28)$$

with the corresponding non-dimensional conservation law (after dropping the tilde)

$$\frac{\partial \mathcal{M}}{\partial t} + Fr \left(\frac{\partial}{\partial x}, \frac{1}{\epsilon} \frac{\partial}{\partial y} \right) \cdot (\mathbf{u} \cdot \mathcal{M}) + \frac{\partial}{\partial x} \left(\frac{1}{2} (1 + Fr \eta)^2 \right) = 0. \quad (2.29)$$

3. Balance model: weakly nonlinear regime

In this section we demonstrate how the ‘slaving method’ proposed by Warn *et al.* (1995) can be used to systematically derive a hierarchy of balance models from the weakly nonlinear equatorial SWEs.

3.1. Slaving method

We begin with a short description of the ‘slaving’ method. Consider a system of differential equations of the form:

$$\frac{\partial \mathbf{s}}{\partial t} = \mathcal{S}(\mathbf{s}, \mathbf{f}; \epsilon) \quad (3.1a)$$

$$\frac{\partial \mathbf{f}}{\partial t} + \frac{\Gamma \mathbf{f}}{\epsilon} = \mathcal{F}(\mathbf{s}, \mathbf{f}; \epsilon), \quad (3.1b)$$

where \mathbf{s} and \mathbf{f} are respectively the slow and fast variables, and \mathcal{S} and \mathcal{F} are nonlinear operators. The linear operator Γ is assumed to be invertible. From a geometrical point of view, the balance motion occurs over a subset of the phase space and is assumed to be constrained onto a geometrical manifold described via $\mathcal{M}(\mathbf{f}, \mathbf{s}; \epsilon) = 0$. The slaving technique assumes that this slow manifold can be expressed in an explicit form

$$\mathbf{f} = U(\mathbf{s}; \epsilon), \quad (3.2)$$

which tacitly views \mathbf{f} as a function of (or ‘slaved’ onto) \mathbf{s} . Substituting the above into (3.1b) results in

$$\mathbf{f} = \epsilon \Gamma^{-1} (\mathcal{F}(\mathbf{s}, \mathbf{f}; \epsilon) - \mathbb{F} \mathcal{S}(\mathbf{s}, \mathbf{f}; \epsilon)), \quad (3.3)$$

with \mathbb{F} being the linearization of U about \mathbf{s} . When $\epsilon \ll 1$, we can make analytical progress via asymptotic methods; but instead of expanding all variables in series of ϵ , only the balance relation is expanded:

$$\mathbf{f} = U(\mathbf{s}; \epsilon) = \mathbf{f}_0 + \epsilon \mathbf{f}_1 + \epsilon^2 \mathbf{f}_2 + \epsilon^3 \mathbf{f}_3 + \dots \quad (3.4)$$

This leads to

$$\frac{\partial \mathbf{f}}{\partial t} = (\mathbb{F}_0 + \epsilon \mathbb{F}_1 + \epsilon^2 \mathbb{F}_2 + \epsilon^3 \mathbb{F}_3 + \dots) \frac{\partial \mathbf{s}}{\partial t}, \quad (3.5)$$

with \mathbb{F}_n being the linearization of \mathbf{f}_n about \mathbf{s} . The terms \mathbf{f}_n in the asymptotic balance relation (3.4) can then be solved iteratively with the aid of (3.1a) and (3.3) as follows. Collecting the $O(1)$ terms in (3.3), we have $\mathbf{f}_0 = 0$. Substituting $\mathbf{f}_0 = 0$ into (3.1a) and collecting the $O(1)$ terms results in a prognostic equation that only involves \mathbf{s} . At the next order, (3.3) becomes $\mathbf{f}_1 = \Gamma^{-1} (\mathcal{F}(\mathbf{s}, \mathbf{f}_0; \epsilon) - \mathbb{F}_0 \mathcal{S}(\mathbf{s}, \mathbf{f}_0; \epsilon)) = \Gamma^{-1} (\mathcal{F}(\mathbf{s}, 0; \epsilon))$, and therefore \mathbf{f}_1 is again expressed solely in terms of \mathbf{s} . This can be repeated

indefinitely, leading to a hierarchy of balance models with increasing accuracy:

<u>Balance relation</u>	<u>Prognostic equation</u>
$O(1) : f = f_0 = 0$	$\frac{\partial \mathbf{s}}{\partial t} = \mathcal{S}(\mathbf{s}, 0; \epsilon) _{O(1)} \quad (3.6a)$
$O(\epsilon) : f = \epsilon \Gamma^{-1} \mathcal{F}(\mathbf{s}, 0; \epsilon) _{O(1)}$	$\frac{\partial \mathbf{s}}{\partial t} = \mathcal{S}(\mathbf{s}, \mathbf{f}; \epsilon) _{O(\epsilon)} \quad (3.6b)$
$O(\epsilon^2) : f = \epsilon \Gamma^{-1} (\mathcal{F}(\mathbf{s}, \mathbf{f}; \epsilon) - \mathbb{F} \mathcal{S}(\mathbf{s}, \mathbf{f}; \epsilon)) _{O(\epsilon)}$	$\frac{\partial \mathbf{s}}{\partial t} = \mathcal{S}(\mathbf{s}, \mathbf{f}; \epsilon) _{O(\epsilon^2)} \quad (3.6c)$
\vdots	\vdots
$O(\epsilon^n) : f = \epsilon \Gamma^{-1} (\mathcal{F}(\mathbf{s}, \mathbf{f}; \epsilon) - \mathbb{F} \mathcal{S}(\mathbf{s}, \mathbf{f}; \epsilon)) _{O(\epsilon^{n-1})}$	$\frac{\partial \mathbf{s}}{\partial t} = \mathcal{S}(\mathbf{s}, \mathbf{f}; \epsilon) _{O(\epsilon^n)}, \quad (3.6d)$

where $|_{O(\epsilon^n)}$ denotes an expansion retaining terms up to $O(\epsilon^n)$. The key is to recognize that the balance relations for \mathbf{f} in (3.6) can be solved explicitly; for example, when computing \mathbf{f} in (3.6d), the quantity on the right-hand side is readily expandable to $O(\epsilon^{n-1})$, since f_0, f_1, \dots, f_{n-1} are already known from the previous orders, and thus the balance relation gives us f_n . With this, we can easily write down the $O(\epsilon^n)$ prognostic equation for the slow variable \mathbf{s} .

3.2. Slaving method applied to equatorial SWEs: weakly nonlinear limit

We first apply the slaving method to the equatorial SWEs, assuming that the nonlinear terms are small by demanding that $Fr = \epsilon$ in (2.21), which implies $u \ll c$. Equation (2.21) can be rewritten as

$$(yu + \eta_y)_t + \frac{1}{\epsilon} \left(\frac{\partial^2}{\partial y^2} - y^2 \right) v = - (u_y + y\eta)_x - yvu_y - (\eta v)_{yy} - \epsilon (yuu_x + (\eta u)_{xy}) \quad (3.7a)$$

$$v_t + \frac{1}{\epsilon} (yu + \eta_y) = - (\epsilon uv_x + vv_y) \quad (3.7b)$$

and

$$\eta_t + u_x + \frac{1}{\epsilon} v_y + \epsilon (\eta u)_x + (\eta v)_y = 0, \quad (3.7c)$$

where (3.7a) is obtained by differentiating (2.21c) with respect to y and adding y times (2.21a). Comparing (3.7a)–(3.7b) with (3.1b), we see that v and $yu + \eta_y$ are identified as the fast variables. The third equation (3.7c) is not of the same form as (3.1a), which suggests that η is not an unequivocally slow variable, yet it is nonetheless possible to treat η as such (Warn *et al.* 1995). The form of the balance relations will then be

$$u = u(\eta; \epsilon) \quad \text{and} \quad v = v(\eta; \epsilon), \quad (3.8)$$

which is consonant with the goal of estimating the wind field from the mass field. We expand the above in a power series of ϵ when $\epsilon \ll 1$:

$$u(\eta) = u_0(\eta) + \epsilon u_1(\eta) + \epsilon^2 u_2(\eta) + \dots, \quad (3.9a)$$

and

$$v(\eta) = v_0(\eta) + \epsilon v_1(\eta) + \epsilon^2 v_2(\eta) + \dots. \quad (3.9b)$$

For conciseness, we will also define two differential operators:

$$\mathcal{L}_1 = \frac{\partial}{\partial y} \frac{1}{y} \frac{\partial}{\partial y} - y \quad \text{and} \quad \mathcal{L}_2 = \frac{\partial^2}{\partial y^2} - y^2. \quad (3.10)$$

Here \mathcal{L}_2 is invertible in the sense that $\mathcal{L}_2\phi = f(y) \implies \phi = \mathcal{L}_2^{-1}f$; in addition, $f(y) = 0 \implies \phi = 0$. The exact definition of \mathcal{L}_2^{-1} is discussed in § A.1.

3.2.1. Leading order: $O(1)$

Substituting (3.9) into (3.7a–b), and collecting the $O(\epsilon^{-1})$ terms, we have

$$\mathcal{L}_2 v_0 = 0 \quad \text{and} \quad y u_0 + \eta_y = 0, \quad (3.11a,b)$$

which leads to

$$u_0 = -\eta_y/y \quad \text{and} \quad v_0 = 0. \quad (3.12a,b)$$

In addition, collecting $O(1)$ terms in (3.7c) results in

$$\eta_t + u_{0x} + v_{1y} = 0. \quad (3.13)$$

To close the $O(1)$ system therefore requires v_1 , which is the $O(\epsilon)$ -correction to v_0 . Collecting the $O(1)$ terms in (3.7a), and noting that $(yu + \eta_y)_t = O(\epsilon)$, we have

$$v_1 = -\mathcal{L}_2^{-1}(u_{0y} + y\eta)_x = \mathcal{L}_2^{-1}\mathcal{L}_1\eta_x, \quad (3.14)$$

which alternatively can be written as

$$\begin{aligned} v_1 &= \mathcal{L}_2^{-1} \left(\frac{\partial}{\partial y} \frac{1}{y} \frac{\partial \eta_x}{\partial y} - y\eta_x \right) = \mathcal{L}_2^{-1} \left(\frac{\partial^2}{\partial y^2} \frac{\eta_x}{y} + \frac{\partial}{\partial y} \frac{2\eta_x}{y} - y\eta_x \right) \\ &= \frac{\eta_x}{y} + \mathcal{L}_2^{-1} \frac{\partial}{\partial y} \frac{2\eta_x}{y}. \end{aligned} \quad (3.15)$$

We can see that v_1 contains both a geostrophic component η_x/y and an ageostrophic component that comes in at the same order. For a Kelvin wave $\eta = \exp(-y^2/2)$, v_1 vanishes identically as expected; this is in contrast with other equatorial balance models (e.g. Saujani 2005; Theiss & Mohebalhojeh 2009; Verkley & van der Velde 2010), which do not diagnose Kelvin waves correctly.

Since u_0 and v_1 are both expressed in terms of η , η is the only dependent variable in the prognostic equation (3.13). To summarize, the $O(1)$ -balance model is given by

$$u = -\frac{1}{y} \frac{\partial \eta}{\partial y}, \quad v = 0 \quad (3.16a,b)$$

and

$$\eta_t + u_{0x} + v_{1y} = \frac{\partial \eta}{\partial t} + \frac{\partial}{\partial x} \left(\frac{\partial}{\partial y} \mathcal{L}_2^{-1} \mathcal{L}_1 \eta - \frac{1}{y} \frac{\partial \eta}{\partial y} \right) = 0. \quad (3.16c)$$

The meridional velocity in the prognostic equation (3.16c) is the next-order correction instead of the meridional velocity given by the balance relation (3.16a,b). This is in contrast to the PV-based QG model, where the velocity field in the prognostic equation is the same as that diagnosed from the balance relations; the difference is due to the fact that η is not unequivocally slow (Warn *et al.* 1995).

3.2.2. Next order: $O(\epsilon)$

The higher-order terms in the expansion can be computed with relative ease. We have already computed v_1 when deriving the $O(1)$ -model, while the $O(1)$ -terms in (3.7b) give

$$u_1 = 0. \quad (3.17)$$

There is thus no $O(\epsilon)$ correction to u . The prognostic equation involves v_2 , which is determined via the $O(\epsilon)$ terms from (3.7a):

$$\mathcal{L}_2 v_2 = - (yu_0 u_{0x} + yv_1 u_{0y} + (\eta u_0)_{xy} + (\eta v_1)_{yy}). \quad (3.18)$$

Note that since $yu_0 + \eta_y$ and u_1 both vanish, $(yu + \eta_y)_t = O(\epsilon^2)$ and thus does not appear in (3.18). As u_0 and v_1 are expressed in terms of η , v_2 is in fact a function of η . From (3.7c) we can also write

$$\eta_t + u_{0x} + (v_1 + \epsilon v_2)_y + \epsilon ((\eta u_0)_x + (\eta v_1)_y) = 0, \quad (3.19)$$

which is again a single prognostic equation for η .

3.2.3. Higher order: $O(\epsilon^2)$

The $O(\epsilon^2)$ correction for u can be found via (3.7b):

$$yu_2 = -v_t|_{O(1)}. \quad (3.20)$$

To compute u_2 , note that

$$\begin{aligned} v_t &= \frac{\partial}{\partial t}(v_0 + \epsilon v_1 + \dots) = \frac{\partial}{\partial x} \mathcal{L}_2^{-1} \mathcal{L}_1 \eta_t + O(\epsilon) \\ &= -\frac{\partial}{\partial x} \mathcal{L}_2^{-1} \mathcal{L}_1 (u_{0x} + v_{1y}) + O(\epsilon), \end{aligned} \quad (3.21)$$

and since $\partial_{yy} \mathcal{L}_2^{-1} = (\partial_{yy} - y^2 + y^2) \mathcal{L}_2^{-1} = 1 + y^2 \mathcal{L}_2^{-1}$, the above can be simplified:

$$\begin{aligned} \mathcal{L}_1 (u_{0x} + v_{1y}) &= \frac{\partial}{\partial x} \left(-y + \frac{\partial}{\partial y} \frac{1}{y} \frac{\partial}{\partial y} \right) \left(-\frac{1}{y} \frac{\partial \eta}{\partial y} + \frac{\partial}{\partial y} \mathcal{L}_2^{-1} \mathcal{L}_1 \eta \right) \\ &= \frac{\partial}{\partial x} \left(\frac{\partial \eta}{\partial y} - y \frac{\partial}{\partial y} \mathcal{L}_2^{-1} \mathcal{L}_1 \eta + \frac{\partial}{\partial y} \frac{1}{y} \left(-\frac{\partial}{\partial y} \frac{1}{y} \frac{\partial \eta}{\partial y} + \frac{\partial^2}{\partial y^2} \mathcal{L}_2^{-1} \mathcal{L}_1 \eta \right) \right) \\ &= \frac{\partial}{\partial x} \left(\frac{\partial \eta}{\partial y} - y \frac{\partial}{\partial y} \mathcal{L}_2^{-1} \mathcal{L}_1 \eta + \frac{\partial}{\partial y} \frac{1}{y} (-y \eta + y^2 \mathcal{L}_2^{-1} \mathcal{L}_1 \eta) \right) \\ &= \mathcal{L}_2^{-1} \mathcal{L}_1 \eta_x. \end{aligned} \quad (3.22)$$

We thus have, from (3.25), (3.22) and (3.21):

$$u_2 = \frac{1}{y} \frac{\partial}{\partial x} \mathcal{L}_2^{-1} \mathcal{L}_1 (u_{0x} + v_{1y}) = \frac{1}{y} \frac{\partial^2}{\partial x^2} \mathcal{L}_2^{-2} \mathcal{L}_1 \eta, \quad (3.23)$$

with \mathcal{L}_2^{-n} being a shorthand for applying \mathcal{L}_2^{-1} n times. As in the previous orders, we need to compute v_3 for the prognostic equation. From the $O(\epsilon^2)$ terms in (3.7a),

$$\mathcal{L}_2 v_3 = -\frac{1}{\epsilon^2} (yu + \eta_y)_t|_{O(\epsilon^2)} - (yv_2 u_{0y} + (\eta v_2)_{yy}) - u_{2xy}. \quad (3.24)$$

Note that

$$\begin{aligned} \frac{\partial}{\partial t} (yu + \eta_y) &= \epsilon^2 y \frac{\partial u_2}{\partial t} + O(\epsilon^3) = \epsilon^2 \frac{\partial^2}{\partial x^2} \mathcal{L}_2^{-2} \mathcal{L}_1 \eta_t + O(\epsilon^3) \\ &= -\epsilon^2 \frac{\partial^2}{\partial x^2} \mathcal{L}_2^{-2} \mathcal{L}_1 (u_{0x} + v_{1y}) + O(\epsilon^3), \end{aligned} \quad (3.25)$$

where the divergence term can again be simplified using (3.22). We can then substitute (3.25) into (3.24) and obtain

$$v_3 = \mathcal{L}_2^{-1} \left[\frac{\partial^3}{\partial x^3} \left(\mathcal{L}_2^{-3} \mathcal{L}_1 \eta - \frac{\partial}{\partial y} \frac{1}{y} \mathcal{L}_2^{-2} \mathcal{L}_1 \eta \right) - y v_2 \frac{\partial u_0}{\partial y} - \frac{\partial^2 (\eta v_2)}{\partial y^2} \right]. \quad (3.26)$$

To summarize, the balance relations including terms up to $O(\epsilon^2)$ are

$$u = (-\eta_y + \epsilon^2 \mathcal{L}_2^{-2} \mathcal{L}_1 \eta_{xx}) / y \quad (3.27a)$$

and

$$v = \epsilon \mathcal{L}_2^{-1} \mathcal{L}_1 \eta_x - \epsilon^2 \mathcal{L}_2^{-1} (y u_0 u_{0x} + y v_1 u_{0y} + (\eta u_0)_{xy} + (\eta v_1)_{yy}), \quad (3.27b)$$

while the corresponding prognostic equation for the balanced motion is

$$\eta_t + (u_0 + \epsilon^2 u_2)_x + (v_1 + \epsilon v_2 + \epsilon^2 v_3)_y + \epsilon (\eta u_0)_x + \epsilon (\eta (v_1 + \epsilon v_2))_y = 0. \quad (3.27c)$$

3.3. Linear waves in the balance model

To demonstrate that the balance models derived in the previous section filter out fast IG and MRG waves, we linearize our model about the resting state, and explore the dynamics in terms of normal modes.

3.3.1. $O(1)$ balance model

The prognostic equation for the leading order balance model is (3.16c). We seek normal mode solutions of the form $\eta = \check{\eta}(y) \exp(i(kx - \omega t))$, which results in an eigenvalue problem for η :

$$\mathcal{L} \check{\eta} \equiv \left(\frac{\partial}{\partial y} \mathcal{L}_2^{-1} \mathcal{L}_1 - \frac{1}{y} \frac{\partial}{\partial y} \right) \check{\eta} = \frac{\omega}{k} \check{\eta}. \quad (3.28)$$

By substitution, we can check that the Kelvin wave solution

$$K = \frac{1}{\sqrt[4]{\pi}} e^{-(1/2)y^2} = \phi_0 \quad \text{with } \omega = k \quad (3.29a)$$

satisfies (3.28) exactly. In addition, we also have Rossby wave solutions given by

$$R_n = \frac{1}{\sqrt{2n+1}} \left(\sqrt{n} \phi_{n+1} + \sqrt{n+1} \phi_{n-1} \right), \quad \omega = -\frac{k}{2n+1} \quad \text{and } n = 1, 2, \dots \quad (3.29b)$$

where $\{\phi_n\}$ is the set of orthonormal Hermite functions. The normalization is chosen such that

$$\langle R_n, R_n \rangle = 1, \quad (3.30)$$

with the inner product defined as

$$\langle F(y), G(y) \rangle = \int_{-\infty}^{\infty} F(y) G(y) dy. \quad (3.31)$$

Since the prognostic equation is first order in time, there can be no more than one wave mode for a given k , and hence IG waves do not exist in the balance model. Furthermore, as (3.28) becomes singular at $y = 0$ unless η_y vanishes identically, this regularity condition also filters out MRG waves as $\eta_y \neq 0$ at the equator for these modes (cf. (2.15)).

3.3.2. $O(\epsilon^2)$ balance model

For the $O(\epsilon)$ -prognostic equation given by (3.19), v_2 is nonlinear in η and vanishes upon linearization; the linearized $O(\epsilon)$ -model is then identical to the $O(1)$ model. The prognostic equation (3.27c) at $O(\epsilon^2)$, with the nonlinear terms dropped and using \mathcal{L} defined in (3.28), is

$$\frac{\partial \eta}{\partial t} + \frac{\partial}{\partial x} \mathcal{L} \eta + \epsilon^2 \frac{\partial^3}{\partial x^3} \left(\frac{\partial}{\partial y} \mathcal{L}_2^{-4} \mathcal{L}_1 \eta - \left(\frac{\partial}{\partial y} \mathcal{L}_2^{-1} \frac{\partial}{\partial y} - 1 \right) \left(\frac{1}{y} \mathcal{L}_2^{-2} \mathcal{L}_1 \eta \right) \right) = 0. \quad (3.32)$$

The corresponding eigenvalue problem for the normal mode solution is

$$\left[\mathcal{L} - \epsilon^2 k^2 \left(\frac{\partial}{\partial y} \mathcal{L}_2^{-4} \mathcal{L}_1 - \left(\frac{\partial}{\partial y} \mathcal{L}_2^{-1} \frac{\partial}{\partial y} - 1 \right) \left(\frac{1}{y} \mathcal{L}_2^{-2} \mathcal{L}_1 \right) \right) \right] \check{\eta} = \frac{\omega}{k} \check{\eta}. \quad (3.33)$$

We now determine how the $O(\epsilon^2)$ terms affect the dispersion relation. Note that $\mathcal{L}_1 \phi_0$ vanishes identically, and thus the Kelvin wave solution $\eta = K = \phi_0$ with dispersion relation $\omega = k$ satisfies the above eigenvalue problem exactly. Exact Rossby wave solutions to the above equation are difficult to find, but we can make analytical progress via an asymptotic expansion:

$$\check{\eta} = \check{\eta}_0 + \epsilon^2 \check{\eta}_2 + \dots \quad (3.34a)$$

and

$$\omega = \omega_0 + \epsilon^2 \omega_2 + \dots. \quad (3.34b)$$

At leading order,

$$\mathcal{L} \check{\eta}_0 = \frac{\omega_0}{k} \check{\eta}_0, \quad (3.35)$$

which is exactly the same eigenvalue problem we have dealt with for the leading-order balance model (cf. (3.28)). We thus take

$$\check{\eta}_0 = R_n \quad \text{and} \quad \omega_0 = -\frac{k}{2n+1}. \quad (3.36)$$

At $O(\epsilon^2)$:

$$\left(\mathcal{L} - \frac{\omega_0}{k} \right) \check{\eta}_2 = \left[\frac{\omega_2}{k} + k^2 \left(\frac{\partial}{\partial y} \mathcal{L}_2^{-4} \mathcal{L}_1 - \left(\frac{\partial}{\partial y} \mathcal{L}_2^{-1} \frac{\partial}{\partial y} - 1 \right) \left(\frac{1}{y} \mathcal{L}_2^{-2} \mathcal{L}_1 \right) \right) \right] \check{\eta}_0. \quad (3.37)$$

The correction to the dispersion relation, ω_2 , can be found via a solvability condition. More specifically, the quantity on the right-hand side of (3.37) must be orthogonal to the homogeneous solution of the adjoint problem to ensure that (3.37) has a unique solution. First note that

$$\begin{aligned} \mathcal{L}_1 &= -y + \frac{\partial}{\partial y} \frac{1}{y} \frac{\partial}{\partial y} = \mathcal{L}_2 \frac{1}{y} - \frac{\partial}{\partial y} \frac{1}{y^2} \\ \implies \mathcal{L} &= \frac{\partial}{\partial y} \mathcal{L}_2^{-1} \mathcal{L}_1 - \frac{1}{y} \frac{\partial}{\partial y} = \left(\frac{\partial}{\partial y} \mathcal{L}_2^{-1} \frac{\partial}{\partial y} - 1 \right) \frac{1}{y^2}. \end{aligned} \quad (3.38)$$

The adjoint operator is obtained via integration by parts:

$$\langle \check{\eta}_0^\dagger, \mathcal{L} \check{\eta}_0 \rangle = \left\langle \check{\eta}_0^\dagger, \left(\frac{\partial}{\partial y} \mathcal{L}_2^{-1} \frac{\partial}{\partial y} - 1 \right) \frac{\check{\eta}_0}{y^2} \right\rangle = \left\langle \frac{1}{y^2} \left(\frac{\partial}{\partial y} \mathcal{L}_2^{-1} \frac{\partial}{\partial y} - 1 \right) \check{\eta}_0^\dagger, \check{\eta}_0 \right\rangle, \quad (3.39)$$

giving us the adjoint operator:

$$\mathcal{L}^\dagger = \frac{1}{y^2} \left(\frac{\partial}{\partial y} \mathcal{L}_2^{-1} \frac{\partial}{\partial y} - 1 \right). \quad (3.40)$$

Dividing (3.35) by y^2 , with \mathcal{L} written as in (3.38), results in

$$\frac{1}{y^2} \left(\frac{\partial}{\partial y} \mathcal{L}_2^{-1} \frac{\partial}{\partial y} - 1 \right) \frac{\check{\eta}_0}{y^2} = \mathcal{L}^\dagger \frac{\check{\eta}_0}{y^2} = \frac{\omega_0}{k} \frac{\check{\eta}_0}{y^2}; \quad (3.41)$$

therefore $\check{\eta}_0/y^2$ is the homogeneous solution to the adjoint problem. We are now in a position to derive the solvability condition. Multiplying (3.37) by $\check{\eta}_0/y^2$ and integrating from $-\infty$ to ∞ , the left-hand side vanishes and leaves us with

$$\frac{\omega_2}{k} \left\langle \frac{\check{\eta}_0}{y^2}, \check{\eta}_0 \right\rangle = k^2 \frac{2\sqrt{2n(n+1)}}{(2n+1)^{9/2}} \left(\left\langle \frac{\check{\eta}_0}{y^2}, \frac{\partial \phi_n}{\partial y} \right\rangle - (2n+1)^2 \left\langle \frac{\check{\eta}_0}{y^2}, y^2 \mathcal{L}^\dagger \frac{\phi_n}{y} \right\rangle \right). \quad (3.42)$$

To arrive at the above expression we have utilized the following relations:

$$\mathcal{L}_1 \check{\eta}_0 = \mathcal{L}_1 R_n = -\frac{2\sqrt{2n(n+1)}}{\sqrt{2n+1}} \phi_n \quad \text{and} \quad \mathcal{L}_2^{-1} \phi_n = -\frac{1}{2n+1} \phi_n. \quad (3.43)$$

The second inner product on the right-hand side of (3.42) can be further simplified, using integration by parts and (3.35):

$$\left\langle \frac{\check{\eta}_0}{y^2}, y^2 \mathcal{L}^\dagger \frac{\phi_n}{y} \right\rangle = \left\langle \mathcal{L} \check{\eta}_0, \frac{\phi_n}{y} \right\rangle = -\frac{1}{2n+1} \left\langle \check{\eta}_0, \frac{\phi_n}{y} \right\rangle = -\frac{\omega_0}{k} \left\langle \frac{\check{\eta}_0}{y^2}, y \phi_n \right\rangle. \quad (3.44)$$

Substituting the above back into (3.42), the two inner products can then be combined using the definition of $R_n = \check{\eta}_0$:

$$\left\langle \frac{\check{\eta}_0}{y^2}, \phi_n' + (2n+1)y\phi_n \right\rangle = \sqrt{2n(n+1)(2n+1)} \left\langle \frac{\check{\eta}_0}{y^2}, \check{\eta}_0 \right\rangle, \quad (3.45)$$

and (3.42) simplifies to

$$\frac{\omega_2}{k} \left\langle \frac{\check{\eta}_0}{y^2}, \check{\eta}_0 \right\rangle = k^2 \frac{4n(n+1)}{(2n+1)^4} \left\langle \frac{\check{\eta}_0}{y^2}, \check{\eta}_0 \right\rangle \implies \omega_2 = k^3 \frac{4n(n+1)}{(2n+1)^4}. \quad (3.46)$$

Then ω_2 can be substituted back into (3.37), and a particular solution is

$$\check{\eta}_2 = \frac{2\sqrt{2n(n+1)}}{(2n+1)^{5/2}} k^2 y \phi_n. \quad (3.47)$$

To summarize, the asymptotic Rossby wave solution to the $O(\epsilon^2)$ prognostic equation is:

$$\check{\eta} = \frac{1}{\sqrt{2n+1}} \left(\sqrt{n} \phi_{n+1} + \sqrt{n+1} \phi_{n-1} + \epsilon^2 k^2 \frac{2\sqrt{2n(n+1)}}{(2n+1)^2} y \phi_n \right) + O(\epsilon^4), \quad (3.48a)$$

with the dispersion relation given by

$$\omega = -\frac{k}{2n+1} + \epsilon^2 k^3 \frac{4n(n+1)}{(2n+1)^4} + O(\epsilon^4). \quad (3.48b)$$

Figure 2(a) is a comparison between the dispersion relation for the equatorial Rossby waves in the linear SWEs and the two balance models discussed in this section

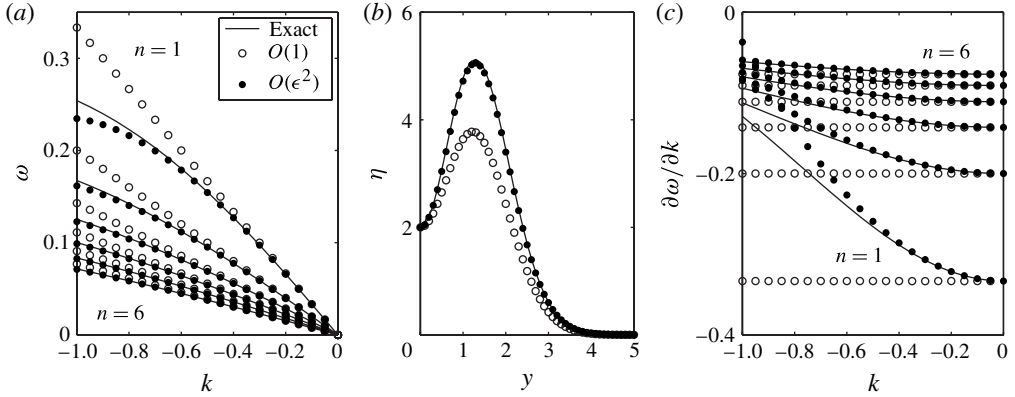


FIGURE 2. (a) Dispersion relation for the equatorial Rossby waves in the SWEs (solid line), the leading-order balance model (open circles) and $O(\epsilon^2)$ -balance model (dots). The scale for the frequency ω is $\sqrt{\beta c}$, while the scale for the zonal wavenumber k is $\sqrt{\beta/c}$. Only the first six Rossby wave modes are shown. (b) The $n = 1$ Rossby wave mode for the SWEs and balance models with $\epsilon = 1$ and $k = -1$. (c) The group velocity $\partial\omega/\partial k$ for the first six Rossby wave modes.

(cf. (3.29b) and (3.48b)). Note that we have reverted back to the regular, isotropic scaling by setting $\epsilon = 1$ to allow for proper comparison. While the leading-order balance relations provide good approximation up to $k \approx 0.3$, the higher-order terms can significantly improve the accuracy for higher wavenumbers. The correction is particularly important for the gravest modes. Figure 2(b) is a comparison between η for the exact normal mode solution to the first Rossby wave and the corresponding normal mode in the $O(1)$ - and $O(\epsilon^2)$ -linear balance models for $\epsilon = 1$ and $k = 1$. With the $O(\epsilon^2)$ -terms, the balance model reproduces the Rossby mode quite well even for the isotropic limit.

It is also useful to consider the group velocity as it determines the speed at which wave packets propagate. As Kelvin waves are represented exactly in the balance models we only consider Rossby waves here. From the dispersion relations (2.11), (3.36), and (3.48b), the group velocity $\partial\omega/\partial k$ can be computed:

$$\text{Exact: } \frac{\partial\omega}{\partial k} = \frac{\omega + 2k\omega^2}{2\omega^3 + k} \quad (3.49a)$$

$$O(1): \frac{\partial\omega}{\partial k} = -\frac{1}{2n+1} \quad (3.49b)$$

$$O(\epsilon^2): \frac{\partial\omega}{\partial k} = -\frac{1}{2n+1} + \frac{12n(n+1)}{(2n+1)^4}k^2, \quad (3.49c)$$

where ϵ is set to unity. The group velocities are plotted in figure 2(c). Unlike the SWE model, the Rossby waves in the $O(1)$ -balance model are not dispersive. On the other hand, the higher-order terms in the $O(\epsilon^2)$ -model introduce the dispersive effects, and bring the group velocity of the Rossby waves much closer to that of the full model; however, it should be noted that for the first Rossby wave mode the group velocity diverges significantly as $k \rightarrow -1$.

3.4. Summary

In this section, we have demonstrated that the ‘slaving’ method proposed by Warn *et al.* (1995), with η as the slow variable, can be used to derive equatorial balance models which filter out IG and MRG waves while retaining the slow Kelvin and Rossby waves. The method allows for higher-order terms to be included systematically, which results in more accurate representation of the slow wave modes in the balance model.

4. Balance relations

We now turn our attention to the balance relations. A difficulty arises as a singularity develops in the balance relations due to the Coriolis term vanishing at the equator. Therefore, η must satisfy a regularity condition in order for the balance relations to produce a smooth wind field. This is the focus of §4.1. In §4.2 we demonstrate the accuracy of the linearized balance relations by comparing the wind field reconstructed through the balance relations to the exact normal mode solution. As a final test, the weakly nonlinear balance is used to initialize a numerical integration of the SWEs in §4.3 to demonstrate that it is indeed useful in suppressing fast oscillations.

4.1. Regularity conditions

The issue of regularity arises even in the leading-order model: the zonal wind inverted via zonal geostrophic balance is $u_0 = -\eta_y/y$, which will remain bounded near the equator only if $\eta_y \rightarrow 0$ as $y \rightarrow 0$. In a practical application, the observed η will generally not satisfy the regularity conditions for the balance model, even in the absence of observation errors: one can check that the derivative of the exact Rossby wave solution (cf (2.12b)) does not vanish at the equator whenever n is even. In these cases, if we substitute in the exact expression for η we will wind up with a spurious singularity in the wind fields. Therefore, η has to be modified to satisfy some regularity conditions.

One of the ways to achieve this is to project η onto the normal mode solutions for the balance model. For example, the normal mode solutions, K and R_n , for the $O(1)$ -model given by (3.29) have derivatives that vanish at the equator, and thus the wind field calculated using these modes will be smooth. This is not surprising as we would expect the balance models to behave nicely near the equator. The projection is however complicated by the fact that $\{K, R_n\}$ do not form an orthonormal basis. For higher-order balance models, an additional difficulty is that normal modes are not known exactly, thus a direct projection is not a feasible method to ensure regularity.

In this section we propose a remedy for the aforementioned problems. The first problem can be avoided by first projecting η onto the set of Hermite functions $\{\phi_n\}$, which is a set of basis functions that is orthonormal and complete, before a change of basis to $\{K, R_n\}$. This procedure results in a residual term, which can then be tuned so that the singular terms are small and can be disregarded. In this case we are seeking an approximate, rather than an exact, regularity condition.

4.1.1. $O(1)$ -balance relations

As the set of Hermite functions $\{\phi_n\}$ is orthonormal and complete, we can approximate η as

$$\eta \approx \sum_0^{N+1} \hat{\eta}_n \phi_n \quad \text{with } \hat{\eta}_n = \langle \eta, \phi_n \rangle. \quad (4.1)$$

We can then apply a change of basis to $\{K, R_n\}$:

$$\sum_0^{N+1} \hat{\eta}_n \phi_n = \hat{\eta}_K K + \sum_1^N \hat{\eta}_{R,n} R_n + \hat{\eta}_r \phi_1. \quad (4.2)$$

The last term is the residual from the change of basis from $\{\phi_n\}$ to $\{K, R_n\}$. Given coefficients $\{\hat{\eta}_n\}$, the coefficients $\{\hat{\eta}_K, \hat{\eta}_{R,n}, \hat{\eta}_r\}$ can be determined uniquely: multiplying ϕ_m with $m = 0, 1, 2, \dots, N + 1$ to both sides of (4.2), and integrating from $-\infty$ to ∞ , we obtain $N + 2$ equations. The equations can be written as a matrix equation, where the matrix is upper triangular with non-zero diagonal elements; its determinant, which is equal to the product of the diagonal elements, is also non-zero. The matrix is therefore invertible, and $\{\hat{\eta}_K, \hat{\eta}_{R,n}, \hat{\eta}_r\}$ can be uniquely determined when $\{\hat{\eta}_n\}$ is given:

$$\begin{bmatrix} \hat{\eta}_K \\ \hat{\eta}_r \\ \hat{\eta}_{R,1} \\ \hat{\eta}_{R,2} \\ \vdots \\ \hat{\eta}_{R,N-1} \\ \hat{\eta}_{R,N} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \sqrt{\frac{2}{3}} & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \sqrt{\frac{3}{5}} & \dots & 0 & 0 \\ 0 & 0 & \sqrt{\frac{1}{3}} & 0 & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \sqrt{\frac{2}{5}} & \ddots & \sqrt{\frac{N-2}{2N-5}} & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 & \sqrt{\frac{N-1}{2N-3}} \\ 0 & 0 & 0 & 0 & \dots & \sqrt{\frac{N-1}{2N-1}} & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 & \sqrt{\frac{N}{2N+1}} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\eta}_0 \\ \hat{\eta}_1 \\ \hat{\eta}_2 \\ \hat{\eta}_3 \\ \vdots \\ \hat{\eta}_N \\ \hat{\eta}_{N+1} \end{bmatrix}. \quad (4.3)$$

The advantage of working with the second expression in (4.2) is that the y -derivatives of K and R_n vanish at $y = 0$, and thus information about η_y is contained in the residual term. Using the recurrence relations for Hermite functions in appendix A, one can show that $R'_n = y(\sqrt{n+1}\phi_{n-1} - \sqrt{n}\phi_{n+1})/\sqrt{2n+1}$. Together with $K' = \phi'_0 = -y\phi_0$ and $\phi'_1 = \sqrt{2}\phi_0 - y\phi_1$, substituting (4.2) into (3.12a) yields

$$u_0 = \hat{\eta}_K \phi_0 - \sum_1^N \hat{\eta}_{R,n} \frac{\sqrt{n+1}\phi_{n-1} - \sqrt{n}\phi_{n+1}}{\sqrt{2n+1}} - \hat{\eta}_r \left(\frac{\sqrt{2}\phi_0}{y} - \phi_1 \right). \quad (4.4)$$

If u is to remain finite as $y \rightarrow 0$, $\hat{\eta}_r = 0$. Physically this corresponds to filtering out the MRG wave mode as it has a non-vanishing first derivative at the equator. With $\hat{\eta}_r = 0$ we can ensure that u_0 remains smooth across the equator.

4.1.2. $O(\epsilon)$ -balance relations

u at $O(\epsilon)$ is identical to the leading order and thus the adjustment discussed in the previous section applies directly. Here u_0 can be used to calculate v_1 via (3.14):

$$v_1 = -\mathcal{L}_2^{-1}(u_{0y} + y\eta)_x. \quad (4.5)$$

4.1.3. $O(\epsilon^2)$ -balance relations

At $O(\epsilon^2)$, u is obtained from η via (3.27a). First note that

$$\mathcal{L}_1 R_n = -2\sqrt{\frac{2n(n+1)}{2n+1}}\phi_n \quad \text{and} \quad \mathcal{L}_1 K = 0, \quad (4.6)$$

and thus substituting (4.2) into (3.27a) results in

$$\begin{aligned} u = & \hat{\eta}_K \phi_0 - \sum_1^N \frac{\hat{\eta}_{R,n}}{\sqrt{2n+1}} \left(\sqrt{n+1}\phi_{n-1} - \sqrt{n}\phi_{n+1} \right) - \hat{\eta}_r \left(\frac{\sqrt{2}\phi_0}{y} - \phi_1 \right) \\ & + \frac{\epsilon^2 k^2}{y} \left(\sum_1^N \hat{\eta}_{R,n} \frac{2}{(2n+1)^2} \sqrt{\frac{2n(n+1)}{2n+1}} \phi_n - \hat{\eta}_r \mathcal{L}_2^{-1} \mathcal{L}_2^{-1} \mathcal{L}_1 \phi_1 \right). \end{aligned} \quad (4.7)$$

Choosing

$$\hat{\eta}_r = \epsilon^2 k^2 \sum_1^N \hat{\eta}_{R,n} \frac{2}{(2n+1)^2} \sqrt{\frac{n(n+1)}{2n+1}} \phi_n(0), \quad (4.8)$$

the last term in the second line of (4.7) is then of $O(\epsilon^4)$, and therefore can be neglected without a loss in formal accuracy since the balance model is only accurate up to $O(\epsilon^2)$. Equation (4.7) becomes

$$\begin{aligned} u = & u_0 + \epsilon^2 u_2 = \hat{\eta}_K \phi_0 - \sum_1^N \frac{\hat{\eta}_{R,n}}{\sqrt{2n+1}} \left(\sqrt{n+1}\phi_{n-1} - \sqrt{n}\phi_{n+1} \right) \\ & + \frac{\epsilon^2 k^2}{y} \left(\sum_1^N \hat{\eta}_{R,n} \frac{2}{(2n+1)^2} \sqrt{\frac{n(n+1)}{2n+1}} \left(\sqrt{2}\phi_n - \sqrt{2}\phi_n(0)\phi_0 + y\phi_1 \right) \right). \end{aligned} \quad (4.9)$$

The sum in the second line of (4.9) vanishes at the equator (since $\phi_0(0) = 1$), and thus the $O(\epsilon^2)$ term remains finite. Note that this adjustment only affects the odd component of η , and no adjustment is made when η is even.

At $O(\epsilon^2)$, v is formally given by

$$v = -\epsilon v_1 - \epsilon^2 \mathcal{L}_2^{-1} (y u_0 u_{0x} + y v_1 u_{0y} + (\eta u_0)_{xy} + (\eta v_1)_{yy}), \quad (4.10)$$

with $u_0 = -\eta_y/y$ and $v_1 = \mathcal{L}_2^{-1} \mathcal{L}_1 \eta_x$. To ensure that these two expressions remain smooth one should substitute in η with $\hat{\eta}_r$ set to be zero.

The caveat of handling the regularity in the manner described in this section is that the truncated $O(\epsilon^4)$ term in (4.9) can become unbounded near $y = 0$; however, this singularity is not expected to be physical for a balanced flow, thus justifying the truncation.

4.2. Numerical verification

The accuracy of the linear balance relations can be tested numerically using the linear normal mode solutions. In particular, we would like to demonstrate that the regularization method we have proposed in the previous section does not affect the accuracy of the height inversion (i.e. calculating \mathbf{u} from η using the balance relations). Treating η from a normal mode solution as an observation, we can then infer \mathbf{u}_{inv} from the balance relations and compare \mathbf{u}_{inv} to the exact solution \mathbf{u} , which serves as the ‘truth’.

4.2.1. Method

We utilize the pseudo-spectral method with Hermite functions as the orthogonal basis functions for approximating the functions. Instead of demanding the residual to be orthogonal to the basis set as in the Galerkin spectral method, the pseudo-spectral method demands the residual to be zero at specific collocation points:

$$\eta \approx \sum_0^{N+1} \hat{\eta}_n \phi_n, \quad \text{with } \eta(y_i) = \sum_0^{N+1} \hat{\eta}_n \phi_n(y_i) \quad \text{for } i = 0, 1, 2, \dots, N+1, \quad (4.11)$$

where the set of collocation points $\{y_i\}$ consists of the roots of the $N+2$ th-degree Hermite polynomial. With this approximation, the derivatives can be calculated easily by multiplying $\eta(y_i)$ by a ‘differentiation matrix’. The MATLAB routine for calculating the collocation points and the differentiation matrix is obtained from Weideman & Reddy (2000) (available at <http://dip.sun.ac.za/~weideman/research/differ.html>).

Given the value of η at the collocation points, it is straightforward to calculate the coefficients $\hat{\eta}_n$ in (4.11) as $\phi_n(y_i)$ are known. We can then apply a change of basis using (4.3), and proceed with the calculations outlined in the previous section to calculate u_0, v_0 , etc.

4.2.2. Results

To test the accuracy of the balance relations, we define the error to be given by

$$\text{Error} = \max_{y \in (-\infty, \infty)} |\mathbf{u} - \mathbf{u}_{inv}|. \quad (4.12)$$

Based on the fact that the neglected terms in deriving the $O(1)$ -balance model are of $O(\epsilon)$, the error defined by (4.12) is expected to decrease as ϵ for the wind fields. For the $O(\epsilon)$ -model the error is expected to decrease as ϵ^2 , and so forth.

Here ϵ is varied from 10^{-3} to 1 with $k = \epsilon$, and the normal mode solutions (\mathbf{u}, η) can be calculated using (2.11) and (2.12). Then η is used in the balance relations to calculate the inverted wind field \mathbf{u}_{inv} . We only show the results here for the second Rossby wave mode $n = 2$, but the results are qualitatively similar for several other Rossby modes tested. This mode is chosen because it is asymmetric about the equator, and thus the adjustment outlined in the previous section is applied. We omit Kelvin waves, as they are represented exactly in the linear model.

On the left-hand side of figure 3, the errors for the balance relations are plotted together with the convergence expected based on the order of the balance model, which are in good agreement even for ϵ as large as unity. In addition, the higher-order balance relations have significantly smaller errors for a given ϵ , confirming that the higher-order model provides a better representation of the slow equatorial waves. More importantly, the error in the inverted wind field is still relatively small at $\epsilon = 1$; this is evident from a comparison between the normal mode solutions and the inverted wind field. From figure 3(b,c), it is clear that at $O(\epsilon^2)$, the inverted wind field remains extremely accurate even for $\epsilon = 1$. The maximum error in magnitude of the wind speed is 0.43, which corresponds to an error of 2.9%. We can therefore expect the balance relations to remain useful as we approach the isotropic (i.e. $L_x = L_y$) limit.

4.3. Balance relations and model initialization

The analysis in the previous section cannot be repeated when the nonlinear terms are included, as the corresponding exact solutions are not available; however, alternative testing can be done via model initialization. The idea is that if the state of the shallow-water system is initially on the slow manifold, then as we integrate forward in time

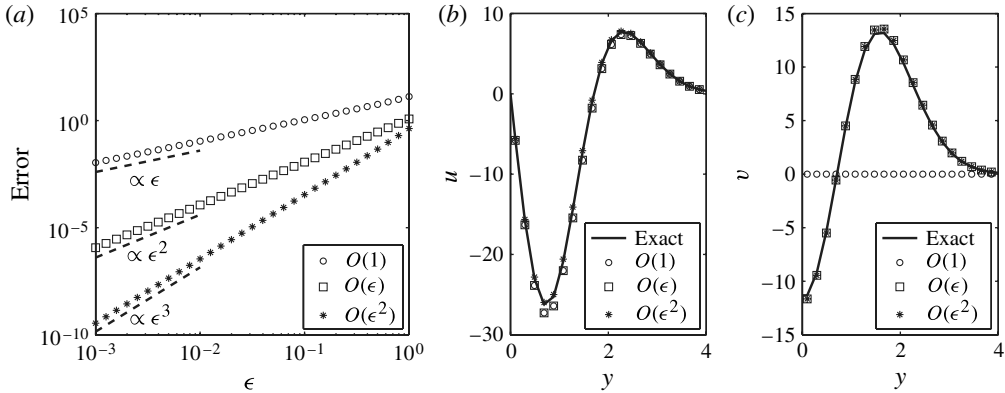


FIGURE 3. (a) Error of the balance relations as a function of ϵ for the second Rossby wave mode. The dashed lines on the left indicate the expected slope for various orders of convergence. The calculations are done with 128 collocation points ($N = 126$). (b) A comparison between the exact solution for u , and the inverted zonal wind via the balance relations. Notice that the $O(1)$ - and $O(\epsilon)$ -balance relations for u are identical for the linear model. $\epsilon = 1$. Here u is an odd function. (c) Same as (b) but for v . Here v inverted from η is identical for the $O(\epsilon)$ - and $O(\epsilon^2)$ -model and is an even function.

the evolution will proceed based on the slow dynamics and free of the high-frequency noise due to the fast dynamics. As the balance relations are just approximations to the slow manifold, in practice the fast component can only be minimized for a given initial state; however, it is expected that the higher-order balance relation will result in less noise (see, for example, Temperton & Williamson 1981; Williamson & Temperton 1981).

In this section we carry out a simple model initialization test. We assume that we initially have observations of η at every grid point, and proceed to use the balance relations (with the nonlinear terms) to calculate the wind field assuming that η describes a balanced flow. This is analogous to the geopotential constrained initialization discussed by Daley (1993). The data is then used as the initial condition for a forward integration in time using the full set of SWEs.

The SWEs are discretized spatially using the pseudo-spectral method with Fourier and Hermite basis as the interpolant respectively for x and y . The time discretization is done using the fourth-order Adams–Bashforth, with fourth-order Runge–Kutta for the initial start up. Here η at $t = 0$ is taken to be a superposition of a Kelvin wave and the first Rossby wave mode, with their amplitudes scaled to satisfy the scaling assumption of $Fr = \epsilon = 0.1$.

The evolution of (u, v, η) at a particular grid point is shown in figure 4 (the behaviour at other points is qualitatively similar). As t has been non-dimensionalized via $1/\sqrt{\beta c}$, the time is measured in terms of inertial periods. We can see that when the initialization is done via the $O(1)$ -balance relations, the variables clearly evolve on two time scales. The fast oscillations in figure 4 occur over several inertial periods, and they can therefore be identified as IG waves. On the other hand, the variables also exhibit changes that take place over much longer periods, and these can be identified as the slow component associated with the slow Kelvin and Rossby wave modes. We can see clearly that the amplitude of the fast oscillation is drastically reduced as we use higher-order balance relations for the initialization, demonstrating the usefulness of

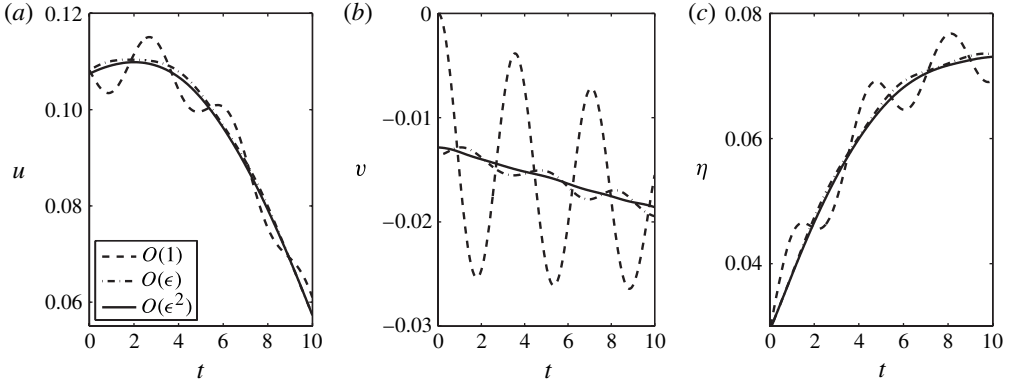


FIGURE 4. Evolution of the shallow-water system using different initializations: (a) u ; (b) v ; (c) η . See the text for details.

the balance relations. In particular these simple experiments suggest that the balance relations correctly diagnose Kelvin waves.

5. Extension to fully nonlinear SWEs

In the previous two sections we have derived balance models for the equatorial SWEs in the weakly nonlinear, large zonal scale regime, and explored their properties. The weakly nonlinear assumption ensures that the leading-order dynamics is linear, but is in fact not necessary. In this section we explore the consequences of allowing Fr to approach unity. The motivation is that atmospheric observations suggest that convectively coupled equatorial waves have equivalent depths of 12–50 m (Wheeler & Kiladis 1999), which corresponds to $c \sim O(10 \text{ m s}^{-1})$; therefore $Fr = O(1)$ is potentially of interest.

Setting $Fr = 1$ while retaining the assumption that $\alpha = \epsilon \ll 1$ in (2.21), we have

$$u_t + uu_x + \frac{1}{\epsilon}(u_y - y)v + h_x = 0, \quad (5.1a)$$

$$v_t + uv_x + \frac{1}{\epsilon}vv_y + \frac{1}{\epsilon}(yu + h_y) = 0 \quad (5.1b)$$

and

$$h_t + (hu)_x + \frac{1}{\epsilon}(hv)_y = 0, \quad (5.1c)$$

where $h = 1 + \eta$. Similar to the approach taken for the weakly nonlinear limit, we combine (5.1a) and (5.1c) to arrive at

$$-(yu + h_y)_t + \frac{1}{\epsilon} \left(-\frac{\partial^2}{\partial y^2} + \frac{y(y - u_y)}{h} \right) hv = (hu)_{xy} + yh_x + yuu_x. \quad (5.1d)$$

5.1. Leading order

We again expand u and v in terms of a power series in ϵ , and use h as the slaving variable. At leading order equations (5.1b)–(5.1d) yield

$$v_0 v_{0y} + yu_0 + h_y = 0, \quad (5.2a)$$

$$\mathcal{L}(hv_0) \equiv \left(-\frac{\partial^2}{\partial y^2} + \frac{y(y - u_{0y})}{h} \right) hv_0 = 0, \quad (5.2b)$$

and

$$h_t + (hu_0)_x + (hv_1)_y = -\frac{1}{\epsilon} (hv_0)_y. \quad (5.2c)$$

Note that \mathcal{L} is different from the similar operator defined for the weakly nonlinear model. At first glance (5.2a–b) are coupled for u_0 and v_0 but they can in fact be solved independently. The slaving method assumes u_0 to be a function of h , and when h is given, the differential operator can be written as $\mathcal{L} = -\partial_{yy}^2 + r(y)$, where $r(y) = y(y - u_{0y})/h$.

Eigenvalue problems of the form

$$-\frac{d^2\Phi}{dy^2} + r(y)\Phi = \lambda\Phi \quad \text{with } \Phi(\pm\infty) = 0 \quad (5.3)$$

represent a singular Sturm–Liouville problem as the domain is unbounded. In the case where $r(y) \rightarrow \infty$ as $y \rightarrow \pm\infty$, the spectrum is known to be discrete and bounded from below (Titchmarsh 1962; Levitan 1988); in other words, there exists an increasing sequence of eigenvalues $\lambda_0 < \lambda_1 < \lambda_2 < \lambda_3 \dots$. Suppose that Φ_0 is the eigenfunction associated with λ_0 , then the Rayleigh quotient gives

$$\lambda_0 = \frac{\int_{-\infty}^{\infty} [\Phi_0'^2 + r(y)\Phi_0^2] dy}{\int_{-\infty}^{\infty} \Phi_0^2 dy}. \quad (5.4)$$

In the case where $r(y)$ is strictly positive, all of the terms in the above integral are positive and λ_0 is guaranteed to be positive. Most importantly, 0 is not an eigenvalue, and thus Fredholm's alternative theorem guarantees a unique solution to the inhomogeneous boundary value problem

$$-\frac{d^2\Phi}{dy^2} + r(y)\Phi = F(y) \quad \text{with } \Phi(\pm\infty) = 0. \quad (5.5)$$

In this case the solution can be expressed in terms of the orthonormal set of eigenfunctions:

$$\Phi = \sum_0^{\infty} \frac{1}{\lambda_n} \langle F, \Phi_n \rangle \Phi_n = \mathcal{L}^{-1}F. \quad (5.6)$$

The above demonstrates that the differential operator is invertible, and in the case where $F(y) = 0$, the only solution is $\Phi = 0$.

Returning to our case where $r(y) = y(y - u_{0y})/h$, the above theorem applies since $r(y)$ approaches y^2 sufficiently far away from the equator (since we would expect $h \rightarrow 1$ and $u_0 \rightarrow 0$ as $y \rightarrow \infty$). Interestingly, the criterion $r(y) > 0$ is satisfied whenever $y(y - u_{0y}) > 0$, which is the stability criterion for a flow to be inertially stable (Ripa 1983). The interpretation is that for an inertially stable flow, the operator \mathcal{L} is invertible and results in a unique solution. From (5.2b),

$$\mathcal{L}(hv_0) = 0 \implies v_0 = 0, \quad (5.7)$$

in which case (5.2a) simplifies to

$$u_0 = -h_y/y. \quad (5.8)$$

Thus at leading order, the balanced flow is again characterized by zonal geostrophic balance and vanishing meridional wind. The prognostic equation requires v_1 , which can be determined from the $O(1)$ terms in (5.1d):

$$\mathcal{L}(hv_1) = (hu_0)_{xy} + yh_x + yu_0u_{0x} \implies v_1 = \frac{1}{h} \mathcal{L}^{-1}((hu_0)_{xy} + yh_x + yu_0u_{0x}), \quad (5.9)$$

and the $O(1)$ -balance model is

$$u = u_0 = -h_y/y, \quad v = 0, \quad (5.10a,b)$$

and

$$h_t + (hu_0)_x + (hv_1)_y = 0. \quad (5.10c)$$

5.2. Higher order: $O(\epsilon)$

As in the weakly nonlinear regime, the higher-order corrections are easily obtained. At the next order, we can see from (5.1b) that $u_1 = 0$, while v_2 , which is needed for the $O(\epsilon)$ -correction to the prognostic equation, is given by

$$\mathcal{L}(hv_2) = 0 \implies v_2 = 0, \quad (5.11)$$

and thus the $O(\epsilon)$ model is given by

$$u = u_0 = -h_y/y \quad (5.12a)$$

$$v = \epsilon v_1 = \frac{\epsilon}{h} \mathcal{L}^{-1}((hu_0)_{xy} + yh_x + yu_0u_{0x}) \quad (5.12b)$$

$$h_t + (hu_0)_x + (hv_1)_y = 0. \quad (5.12c)$$

As (5.12c) is identical to (5.10c), the dynamics of the leading-order model is in fact accurate to $O(\epsilon)$.

It is worth emphasizing that the weakly nonlinear model considered in § 3 is merely a special limit of the nonlinear model. The correspondence at leading order is clear if we revert back to the weakly nonlinear scaling via $(u, v, h) \rightarrow (\epsilon u, \epsilon v, 1 + \epsilon \eta)$. In this case (5.10a,b) becomes identical to (3.16a,b). As for the prognostic equation, note that

$$\mathcal{L} = -\frac{\partial^2}{\partial y^2} + \frac{y(y - \epsilon u_{0y})}{1 + \epsilon \eta} \approx -\mathcal{L}_2 + O(\epsilon) \implies \mathcal{L}^{-1} = -\mathcal{L}_2^{-1} + O(\epsilon). \quad (5.13)$$

See appendix B for a more detailed justification. Expanding the expression for v_1 , (5.12b), results in

$$\begin{aligned} v &= \frac{\epsilon}{1 + \epsilon \eta} \mathcal{L}^{-1}(u_{0xy} + y\eta_x + \epsilon(\eta u_0)_{xy} + \epsilon y u_0 u_{0x}) \\ &= \epsilon \mathcal{L}^{-1}(u_{0xy} + y\eta_x) + O(\epsilon^2) = \epsilon \mathcal{L}_2^{-1} \mathcal{L}_1 \eta_x + O(\epsilon^2). \end{aligned} \quad (5.14)$$

We thus recover $v_1 = \mathcal{L}_2^{-1} \mathcal{L}_1 \eta_x$. Substituting v_1 into (5.10c), and discarding $O(\epsilon)$ and higher terms, we have

$$\eta_t - (\eta_y/y)_x + (\mathcal{L}_2^{-1} \mathcal{L}_1 \eta_x)_y = 0, \quad (5.15)$$

which is identical to (3.16c). The same approach allows us to demonstrate a similar correspondence for the $O(\epsilon^2)$ models.

5.3. Conservation laws

With the slaving method, only the conservation laws associated with the slaving variable are expected to be preserved in the balance model (Warn *et al.* 1995). Our choice of height as the slaving variable therefore ensures that the balance model is mass-conserving. It is nonetheless interesting to determine whether the other conservation laws for the SWEs (see § 2.3) hold. With $Fr = 1$ and expanding u and v in series of ϵ , the three conserved quantities PV, energy and absolute momentum are respectively given by

$$Q = \frac{\epsilon v_x - u_y + y}{h} = \frac{-u_{0y} + y}{h} + O(\epsilon^2), \quad (5.16a)$$

$$\mathcal{E} = h \frac{u^2 + v^2}{2} + \frac{1}{2} h^2 = \frac{1}{2} (hu_0^2 + h^2) + O(\epsilon^2) \quad (5.16b)$$

and

$$\mathcal{M} = h \left(u - \frac{1}{2} y^2 \right) = h \left(u_0 - \frac{1}{2} y^2 \right) + O(\epsilon^2). \quad (5.16c)$$

Denoting the $O(1)$ terms respectively by Q_0 , \mathcal{E}_0 and \mathcal{M}_0 , we now demonstrate that they are conserved by the $O(\epsilon)$ model. We first calculate u_{0t} using (5.12a) and (5.12c):

$$u_{0t} = -\frac{1}{y} \frac{\partial}{\partial y} \frac{\partial \eta}{\partial t} = \frac{1}{y} \left((hu_0)_{xy} + (hv_1)_{yy} \right). \quad (5.17)$$

Note that (5.9) can be rewritten as

$$(hu_0)_{xy} + (hv_1)_{yy} = \frac{y(y - u_{0y})}{h} hv_1 - yh_x - yu_0 u_{0x}. \quad (5.18)$$

Combining (5.17) and (5.18) yields

$$u_{0t} = -u_0 u_{0x} - u_{0y} v_1 + y v_1 - h_x = -u_0 u_{0x} + Q_0 h v_1 - h_x, \quad (5.19)$$

which can also be obtained by collecting the $O(1)$ terms from (5.1a), the zonal momentum equation.

5.3.1. Potential vorticity

Differentiating Q_0 with respect to t , using (5.12c) and (5.19) to respectively eliminate h_t and u_{0t} , and noting that $(u_0 u_{0x})_y = (u_0 u_{0y})_x$, we find

$$\begin{aligned} \frac{\partial Q_0}{\partial t} &= -\frac{u_{0yt}}{h} - \frac{Q_0}{h} \eta_t \\ &= -\frac{1}{h} (Q_0 h v_1)_y + \frac{1}{h} \frac{\partial}{\partial y} (u_0 u_{0x} + h_y) + \frac{Q_0}{h} \left((hu_0)_x + (hv_1)_y \right) \\ &= -v_1 \frac{\partial Q_0}{\partial y} + \frac{1}{h} \frac{\partial}{\partial x} [u_0 (u_{0y} - y)] + \frac{Q_0}{h} (hu_0)_x, \end{aligned} \quad (5.20)$$

where geostrophic balance $h_y = -y u_0$ is used to eliminate h_y . With $u_{0y} - y = -Q_0 h$, $\mathbf{u} = (u_0, v_1)$, and $\nabla = (\partial_x, \partial_y)$, equation (5.20) further simplifies to

$$Q_{0t} + \mathbf{u} \cdot \nabla Q_0 = 0, \quad (5.21)$$

which shows that Q_0 is a material invariant and is conserved following a fluid parcel.

5.3.2. Energy

Differentiating \mathcal{E}_0 with respect to t , we have

$$\frac{\partial}{\partial t} \frac{hu_0^2 + h^2}{2} = \frac{(u_0^2 + 2h)}{2} h_t + hu_0 u_{0t} = - \left(\frac{\mathcal{E}_0}{h} + \frac{h}{2} \right) \nabla \cdot (hu) + hu_0 u_{0t}. \quad (5.22)$$

Using (5.19) and $yu_0 = -h_y$, the last term in the above equation can be written as

$$\begin{aligned} hu_0 u_{0t} &= -(hu_0)(u_0 u_{0x}) - (hv_1)(u_0 u_{0y}) - yu_0 v_1 h - u_0 h h_x \\ &= -hu \cdot \nabla \left(\frac{u_0^2}{2} \right) - hu \cdot \nabla h = -hu \cdot \nabla \left(\frac{\mathcal{E}_0}{h} + \frac{h}{2} \right). \end{aligned} \quad (5.23)$$

Substituting (5.22) into (5.23) then results in

$$\frac{\partial \mathcal{E}_0}{\partial t} + \nabla \cdot \left[\left(\frac{\mathcal{E}_0}{h} + \frac{h^2}{2} \right) \mathbf{u} \right] = 0, \quad (5.24)$$

and therefore the energy \mathcal{E}_0 is an integral invariant and is conserved globally.

5.3.3. Absolute momentum

First observe that (5.19) can also be written as

$$u_{0t} = -\mathbf{u} \cdot \nabla \left(u_0 + \frac{y^2}{2} \right) - h_x = -\mathbf{u} \cdot \nabla \left(\frac{\mathcal{M}_0}{h} \right) - h_x. \quad (5.25)$$

Therefore, \mathcal{M}_{0t} is given by

$$\begin{aligned} \frac{\partial \mathcal{M}_0}{\partial t} &= \left(u_0 - \frac{y^2}{2} \right) h_t + hu_{0t} = - \left(\frac{\mathcal{M}_0}{h} \right) \nabla \cdot (hu) - hu \cdot \nabla \left(\frac{\mathcal{M}_0}{h} \right) - hh_x \\ &= -\nabla \cdot (\mathcal{M}_0 \mathbf{u}) - \frac{\partial h^2}{\partial x} \frac{1}{2}. \end{aligned} \quad (5.26)$$

Equation (5.26) shows that \mathcal{M}_0 is also an integral invariant and a globally conserved quantity.

6. Further extension: Boussinesq equations

The slaving method can be easily extended to a more realistic stratified model for the atmosphere and oceans, such as the Boussinesq equations. The Boussinesq model assumes that the density ρ can be written as a perturbation centred around a constant reference value:

$$\rho = \rho_0 + \rho'(x, y, z, t), \quad (6.1)$$

with $|\rho_0| \gg |\rho'|$. The pressure field can then be decomposed into $p = p_0(z) + p'(x, y, z, t)$, where p_0 is the part that is in hydrostatic balance with the reference density:

$$dp_0/dz = -\rho_0 g. \quad (6.2)$$

Similarly the potential temperature field can be decomposed into a background and perturbed field: $\theta = \theta_0(z) + \theta'(x, y, z, t)$. The equations are

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} + w \frac{\partial u}{\partial z} - \beta y v + \frac{1}{\rho_0} \frac{\partial p'}{\partial x} = 0 \quad (6.3a)$$

$$\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + w \frac{\partial v}{\partial z} + \beta y u + \frac{1}{\rho_0} \frac{\partial p'}{\partial y} = 0 \quad (6.3b)$$

$$\frac{\partial w}{\partial t} + u \frac{\partial w}{\partial x} + v \frac{\partial w}{\partial y} + w \frac{\partial w}{\partial z} - g \frac{\theta'}{\theta_0} + \frac{1}{\rho_0} \frac{\partial p'}{\partial z} = 0 \quad (6.3c)$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0 \quad (6.3d)$$

$$\frac{\partial \theta'}{\partial t} + u \frac{\partial \theta'}{\partial x} + v \frac{\partial \theta'}{\partial y} + w \frac{\partial \theta'}{\partial z} + w \frac{d\theta_0}{dz} = 0. \quad (6.3e)$$

In the case where the vertical scale of motion is small compared with the vertical variation in θ_0 , we can assume θ_0 and $d\theta_0/dz$ to be constants. Defining the Brunt–Väisälä frequency N to be

$$N = \sqrt{\frac{g}{\theta_0} \frac{d\theta_0}{dz}}, \quad (6.4)$$

we proceed to non-dimensionalize the above equations via

$$x = L_x \tilde{x}, \quad y = L_y \tilde{y} = \sqrt{\frac{c}{\beta}} \tilde{y}, \quad z = H \tilde{z}, \quad t = \frac{1}{\epsilon \sqrt{\beta c}} \tilde{t}, \quad (6.5a)$$

$$(u, v) = U (\tilde{u}, \tilde{v}), \quad w = U \frac{H}{L_y} \tilde{w}, \quad p' = Fr \rho_0 c^2 \tilde{p} \quad \text{and} \quad \theta' = Fr \frac{d\theta_0}{dz} H \tilde{\theta}. \quad (6.5b)$$

In the above, the gravity wave speed in a stratified medium is now given by $c = NH$. We again focus on the long-wave limit given by $L_y/L_x = \epsilon \ll 1$. The scale of t is chosen to be $1/\epsilon$ times the time scale for gravity waves $1/\sqrt{\beta c}$ in anticipation of a separation in time scale. The scale for w is chosen to ensure that the vertical and zonal advective time scales are equal. Both p' and θ' are scaled by $Fr = U/c$, which determines the strength of the nonlinearity. Dropping the tildes and defining

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + Fr \left(u \frac{\partial}{\partial x} + \frac{v}{\epsilon} \frac{\partial}{\partial y} + \frac{w}{\epsilon} \frac{\partial}{\partial z} \right), \quad (6.6)$$

the non-dimensionalization reduces (6.3) to

$$\frac{Du}{Dt} - \frac{1}{\epsilon} yv + \frac{\partial p}{\partial x} = 0 \quad (6.7a)$$

$$\frac{Dv}{Dt} + \frac{1}{\epsilon} \left(yu + \frac{\partial p}{\partial y} \right) = 0 \quad (6.7b)$$

$$\frac{Dw}{Dt} + \frac{1}{\epsilon} \frac{L_y^2}{H^2} \left(-\theta + \frac{\partial p}{\partial z} \right) = 0 \quad (6.7c)$$

$$\frac{\partial u}{\partial x} + \frac{1}{\epsilon} \left(\frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right) = 0 \quad (6.7d)$$

$$\frac{D\theta}{Dt} + \frac{1}{\epsilon} w = 0. \quad (6.7e)$$

We would expect that $H \ll L_y$ for most geophysical flows, and thus in the long-wave limit the balanced flow will be strongly hydrostatic. We may then simply take $\theta = \partial p / \partial z$ in lieu of (6.7c). As the hydrostatic Boussinesq equations are isomorphic

to the primitive equations with pressure coordinates (Vallis 2006), the balance model developed here can be easily transferred to the primitive equations.

The form of (6.7) is promising. We already have two diagnostic relations given by hydrostatic balance and the continuity equation (6.7d). From (6.7a) we can see that v should vanish at leading order, while (6.7b) suggests a geostrophic-type balance in the zonal direction. This is similar to the shallow-water system we have dealt with earlier.

We omit the development of the weakly nonlinear model because like its SWEs counterpart, it is subsumed by the fully nonlinear model. However we should point out that the weakly nonlinear balance model admits both Kelvin wave and Rossby wave solutions.

6.1. Fully nonlinear balance model

We directly derive the fully nonlinear balance model by demanding that $Fr = 1$. As in the shallow-water model before, we use the mass field variable θ as the master variable, and assume that the other variables (u, v, w, p) are slaved to θ . Here p can be eliminated in a straightforward manner by differentiating the momentum equations with respect to z . For example, differentiating (6.7b), and using (6.7d) and the hydrostatic approximation one arrives at

$$\frac{Dv_z}{Dt} + u_z v_x - v_z u_x + \frac{1}{\epsilon} \left(y \frac{\partial u}{\partial z} + \frac{\partial \theta}{\partial y} \right) = 0, \quad (6.8)$$

which suggests a thermal wind balance in the zonal direction as expected. Defining $\Gamma = yu_z + \theta_y$, we can similarly use the u -momentum equation and (6.7e), with the aid of the continuity and hydrostatic relations, to write down an equation for Γ :

$$\frac{D\Gamma}{Dt} + \frac{1}{\epsilon} [w_y(\theta_z + 1) + 2\theta_y v_y - y(y - u_y)v_z - vu_z] = \frac{\Gamma v_y}{\epsilon} - u_y \theta_x - y \theta_x. \quad (6.9)$$

Note that since Γ and v are both fast variables and therefore must vanish at leading order, the first term on the right-hand side of (6.9) is in fact at most of $O(\epsilon)$. At this point we have eliminated p , and are left with four equations. The continuity equation suggests that at leading order we have $v_y + w_z = 0$. This motivates the Helmholtz decomposition

$$v = -\psi_z + \chi_y \quad \text{and} \quad w = \psi_y + \chi_z, \quad (6.10a,b)$$

where the stream function ψ and velocity potential χ respectively describe the rotational and divergent motions in the y - z plane. The equations for the Boussinesq system are now in the desired form:

$$\frac{Dv_z}{Dt} + \frac{\Gamma}{\epsilon} = v_z u_x - u_z v_x, \quad (6.11a)$$

$$\frac{1}{\epsilon} \nabla^2 \chi = -u_x, \quad (6.11b)$$

$$\begin{aligned} -\frac{D\Gamma}{Dt} - \frac{1}{\epsilon} \left((\theta_z + 1) \frac{\partial^2}{\partial y^2} - 2\theta_y \frac{\partial^2}{\partial y \partial z} + y(y - u_y) \frac{\partial^2}{\partial z^2} + u_z \frac{\partial}{\partial z} \right) \psi \\ = \frac{1}{\epsilon} [(\theta_z + 1 - y(y - u_y)) \chi_{yz} + 2\theta_y \chi_{yy} - u_z \chi_y] - \frac{\Gamma v_y}{\epsilon} + u_y \theta_x + y \theta_x, \end{aligned} \quad (6.11c)$$

and

$$\frac{D\theta}{Dt} + \frac{w}{\epsilon} = 0, \quad (6.11d)$$

where $\nabla^2 = \partial_{yy} + \partial_{zz}$ denotes the Laplacian in the y - z plane. For simplicity we assume that the domain is periodic in x , and infinite in y and z , with the variables decaying to zero as $|y|$ and $|z| \rightarrow \infty$.

6.1.1. Leading order

Expanding u, v, w, ψ, χ asymptotically, and collecting the leading-order equations, we have

$$\Gamma_0 = yu_{0z} + \theta_y = 0, \quad (6.12a)$$

$$\nabla^2 \chi_0 = 0, \quad (6.12b)$$

$$\begin{aligned} & - \left[(\theta_z + 1) \partial_{yy} - 2\theta_y \partial_{yz} + y(y - u_{0y}) \partial_{zz} + u_{0z} \partial_z \right] \psi_0 \\ & = \left[\theta_z + 1 - y(y - u_{0y}) \right] \chi_{0yz} + 2\theta_y \chi_{0yy} - u_z \chi_{0y} - \Gamma_0 v_{0y}, \end{aligned} \quad (6.12c)$$

and

$$\theta_t + u\theta_x + \frac{1}{\epsilon} \left((v_0 + \epsilon v_1) \theta_y + (w_0 + \epsilon w_1) (\theta_z + 1) \right) = 0. \quad (6.12d)$$

The first three equations allow us to calculate the wind field from θ . The zonal wind is first calculated using the thermal wind balance given by (6.12a). Equation (6.12b) is the Laplace equation with boundary condition $\chi_0 = 0$ at infinity. The standard theory stipulates that χ_0 must vanish identically everywhere. The right-hand side of (6.12c) vanishes since $\Gamma_0 = \chi_0 = 0$, while the differential operator on the left-hand side can be put into divergence form using $yu_{0z} = -\theta_y$:

$$\mathcal{L}\psi_0 \equiv - \left[\frac{\partial}{\partial y} \left((\theta_z + 1) \frac{\partial}{\partial y} - \theta_y \frac{\partial}{\partial z} \right) + \frac{\partial}{\partial z} \left(-\theta_y \frac{\partial}{\partial y} + y(y - u_{0y}) \frac{\partial}{\partial z} \right) \right] \psi_0 = 0. \quad (6.13)$$

Multiplying (6.13) by ψ_0 , integrating over the entire domain, and completing the square results in

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left[((\theta_z + 1)\psi_{0y} - \theta_y \psi_{0z})^2 + ((\theta_z + 1)(y - u_{0y}) - \theta_y^2) \psi_{0z}^2 \right] dy dz = 0. \quad (6.14)$$

The integrand is guaranteed to be positive-definite whenever

$$(\theta_z + 1)(y - u_{0y}) - \theta_y^2 > 0, \quad (6.15)$$

in which case the equality in (6.14) cannot hold unless $\psi_0 = 0$. As with the SWEs, the criterion (6.15) turns out to be identical to the criterion for inertial stability. For a stratified fluid, a flow is inertially stable whenever

$$f^2(1 - Ri^{-1}) - fu_y > 0, \quad (6.16)$$

where the Richardson number $Ri = N^2/u_z^2$ (Andrews, Holton & Leovy 1987). With thermal wind balance, and noting that the potential temperature gradient $(\theta_z + 1)$ is proportional to N^2 , equations (6.15) and (6.16) are identical. The immediate implication is that whenever a zonal flow is inertially stable, $\psi_0 = 0$ is the only solution to satisfy the homogeneous problem $\mathcal{L}\psi_0 = 0$ and vanish at the boundaries. The invertibility of the operator \mathcal{L} is guaranteed under a slightly stronger stability condition: $(\theta_z + 1)(y - u_{0y}) - \theta_y^2 > K$ for some constant $K > 0$. When this condition holds, \mathcal{L} is uniformly elliptic, and the inhomogeneous problem $\mathcal{L}\phi = f$ has a unique solution given by $\phi = \mathcal{L}^{-1}f$ (Evans 2010). For a more detailed discussion on uniformly elliptic operators, see appendix C.

Since $\psi_0 = \chi_0 = 0$, the meridional and vertical wind must vanish at leading order. To complete the leading-order balance model, we must determine the v_1 and w_1 in (6.12d). Collecting the next-order terms from (6.11b) and (6.11c), we have

$$\nabla^2 \chi_1 = -u_{0x}, \quad (6.17a)$$

and

$$-\mathcal{L}\psi_1 = (\theta_z + 1 - y(y - u_{0y}))\chi_{1yz} + 2\theta_y\chi_{1yy} - u_z\chi_{1y} + u_{0y}\theta_x + y\theta_x. \quad (6.17b)$$

Both the Laplace operator and \mathcal{L} are invertible when the flow is inertially stable, thus

$$\chi_1 = -(\nabla^2)^{-1} u_{0x}, \quad (6.18a)$$

and

$$\psi_1 = -\mathcal{L}^{-1} [(\theta_z + 1 - y(y - u_{0y}))\chi_{1yz} + 2\theta_y\chi_{1yy} - u_z\chi_{1y} + u_{0y}\theta_x + y\theta_x]. \quad (6.18b)$$

As u_0 is a function of θ , χ_1 can be computed from (6.18a) and is in turn used to compute ψ_1 using (6.18b). We should emphasize that although the nonlinearity appears at leading order, the diagnostic equations for ψ and χ are linear at all orders. This is in contrast with the Charney–Bolin balance, where the inversion of the velocity field from PV involves solving a nonlinear partial differential equation (PDE).

We should also note that at this point we effectively have the $O(\epsilon)$ -balance model at our disposal: in addition to v_1 and w_1 , the next-order equation obtained from (6.12a) states that $\Gamma_1 = yu_{1z} = 0 \implies u_1 = 0$. Here v_2 and w_2 , which are needed for the prognostic equation, are obtained via

$$\nabla^2 \chi_2 = -u_{1x} = 0, \quad (6.19a)$$

and

$$-\mathcal{L}\psi_2 = (\theta_z + 1 - y(y - u_{0y}))\chi_{2yz} + 2\theta_y\chi_{2yy} - u_z\chi_{2y}. \quad (6.19b)$$

Solving (6.19a) gives $\chi_2 = 0$, which in turn gives $\psi_2 = 0$ as the solution to (6.19b). The conclusion is that the leading-order model is in fact accurate up to $O(\epsilon)$.

To summarize, the $O(\epsilon)$ -balance model is given by the balance relations

$$u = u_0 = - \int (\theta_y/y) dz, \quad v = -\psi_{1z} + \chi_{1y}, \quad w = \psi_{1y} + \chi_{1z}, \quad p = \int \theta dz \quad (6.20a-d)$$

together with the prognostic equation

$$\theta_t + u_0\theta_x + (v_1\theta_y + w_1(\theta_z + 1)) = 0, \quad (6.20e)$$

with ψ_1 and χ_1 given by (6.16). It is worth emphasizing that u_0 , v_1 and w_1 are all functions of θ , and thus the prognostic equation is a closed equation for θ . The derivation for the higher-order model is omitted here as it is largely similar to the weakly nonlinear SWEs.

The quasi-balanced model of Stevens *et al.* (1990) is formally equivalent to the leading-order model developed in this paper. Similar to our linearized model, the linearized quasi-balanced model retains non-dispersive Rossby and Kelvin waves. However, as the quasi-balanced model is derived via a truncation of the meridional momentum equation, it is impossible to systematically include higher-order terms to improve the representation of the Rossby wave dynamics, including dispersive effects (cf. figure 2a); this is in contrast to the asymptotic framework used in the present paper, through which higher-order models can be systematically generated.

7. Discussion and conclusions

In this paper, we have applied the ‘slaving’ method proposed by Warn *et al.* (1995) to derive balance models for the SWEs on the equatorial β -plane, with the goal of eliminating fast IG and MRG waves in the long-wave limit while retaining Rossby and Kelvin waves. Differently from previous approaches, the slow dynamics is described using the mass field variable, h , rather than PV, as equatorial Kelvin waves are invisible in the PV field. The balance models are obtained by expanding the balance relations asymptotically using $L_y/L_x = \epsilon \ll 1$ as the expansion parameter.

We first considered the SWEs in the weakly nonlinear limit $Fr = \epsilon$, where the dynamics of the leading-order system is described through equatorial waves. We demonstrated that the slaving method successfully results in a balance model that retains Rossby and Kelvin waves. Our model is, to the best of our knowledge, the first model to provide a proper diagnostic relation for the zero meridional wind of a Kelvin wave, as well as higher-order corrections that significantly improve the representation of Rossby waves even in the isotropic regime, i.e. as $\epsilon \rightarrow 1$. The accuracy of the balance relations was tested numerically by computing the inverted wind field against the exact solutions, and the error was found to be consistent with the asymptotic theory. Additional tests were carried out using the balance relations to initialize numerical computations; the tests show that the higher-order balance relations significantly reduce the noise due to fast IG waves.

In addition, we have shown that the slaving method can still be applied even when the nonlinearity appears at leading order. This model encompasses the weakly nonlinear regime as a special case, and thus is the general balance model for the long-wave regime. The solvability theorem for the differential operator appearing in the balance relations guarantees that a unique velocity field can always be inverted from the mass field provided the flow is inertially stable. We also applied the slaving method to the Boussinesq equations, demonstrating that analogous balance models can also be derived for stratified models.

The equatorial shallow-water and Boussinesq balance models share several important features. In both cases, the leading-order balance is characterized by geostrophic balance for the zonal wind, while the meridional wind field vanishes. The interpretation is that the geostrophic and ageostrophic components of v are both $O(\epsilon)$, which is in marked contrast with the mid-latitude balance theory. In this regard our theory resembles the mid-latitude semigeostrophic theory, which is also based on anisotropy instead of small Rossby number; however, instead of partitioning v into a geostrophic and an ageostrophic component, v is treated as a single variable. Mathematically it is convenient to decompose the flow in the y - z plane into its rotational and divergent components. The resulting PDEs are both linear, and can be inverted when the flow is inertially stable. The circulation in the y - z plane can therefore be diagnosed entirely from potential temperature.

It is interesting to note that in both the shallow water and Boussinesq models, the horizontal divergence, δ , of the balanced flow is small compared with vorticity ζ . With our non-dimensionalization, the ratio of δ to ζ is given by

$$\frac{\delta}{\zeta} = \frac{\epsilon u_x + v_y}{\epsilon v_x - u_y} = \frac{O(\epsilon)}{O(1)}, \quad (7.1)$$

owing to the fact that v must vanish at leading order for a balanced flow. This appears to be consistent with the scale analysis carried out by Charney (1963), who concluded that large-scale equatorial flow should be horizontally quasi-non-divergent for small

Froude number isotropic flows. In contrast, the quasi-non-divergence in our models is a consequence of the anisotropic scaling, and it may fail as the motion becomes isotropic (i.e. $\epsilon \rightarrow 1$). This clearly occurs for a Kelvin wave in the linear limit, since with $v = 0$, δ/ζ is only small for large zonal scales and becomes $O(1)$ as $\epsilon \rightarrow 1$.

While our weakly nonlinear model successfully reproduces the slow long-wave dynamics, we sacrifice the ability to model the short Rossby wave dynamics; most importantly, the model fails to reproduce the rollover in the dispersion relation for Rossby waves at higher wavenumbers. A natural question is whether it is possible for a balance model to capture Kelvin waves yet accurately describe Rossby wave dynamics at all wavenumbers. In our opinion this is unlikely since the dominant balances in the short- and long-wave regimes are fundamentally different: the slow dynamics in the long-wave regime is characterized by $yu + \eta_y = 0$ and $v = 0$, whereas in the short-wave regime the dominant balance is given by $u_x + v_y = 0$ and $-yv + \eta_x = 0$ (see, for example, Verkley & van der Velde 2010). While the latter ensures that Rossby waves are described accurately in the short-wave limit, it severely distorts the Kelvin waves. It seems likely that in a model where Rossby waves are described accurately at all scales, the Kelvin wave component must be treated separately (e.g. Schubert *et al.* 2009). We should emphasize that for data assimilation it is far more important for the balance relation to accurately describe the slow planetary-scale dynamics, since that is the component that one would hope to capture with the observations.

Another difference between our long-wave model and balance models such as Schubert *et al.* (2009), Theiss & Mohebalhojeh (2009) and Verkley & van der Velde (2010) is that MRG waves are filtered in addition to IG waves, due to the fact that MRG waves are fast in the long-wave regime (see figure 1). Although MRG waves are often associated with convective activity, their contribution to the variance in the equatorial troposphere is small compared with that of Kelvin waves (Wheeler & Kiladis 1999; Roundy & Frank 2004). For the stratosphere, the MRG wave activity is approximately one-sixth of Kelvin wave activity, and thus on balance the inclusion of Kelvin waves outweighs the loss of MRG waves. In any cases, the MRG waves would be difficult to resolve in observations due to limited longitudinal resolution.

We should also note that all of the models considered in this paper assume the motions are adiabatic. As diabatic heating plays an important role in equatorial tropospheric dynamics, it would be interesting to derive diabatically forced balance models. As Gill (1980) has demonstrated, the response to a localized heat source is highly anisotropic due to the equatorial wave guide effect, and therefore the anisotropic balance models developed in this paper are potentially useful as the backbone of a diabatic balance model for planetary-scale motions.

Acknowledgements

This research is funded by the Natural Sciences and Engineering Research Council of Canada. I.C. wishes to acknowledge the financial support from the Government of Ontario. The authors would like to thank the three anonymous reviewers for their helpful comments and suggestions.

Appendix A. Hermite functions

Hermite functions ϕ_n satisfy the differential equation given by

$$\mathcal{L}_2\phi_n = \frac{d^2\phi_n}{dy^2} - y^2\phi_n = -(2n+1)\phi_n \quad \text{with } \phi(\pm\infty) = 0. \quad (\text{A } 1)$$

They are related to n th-degree Hermite polynomials via

$$\phi_n(y) = (2^n n! \sqrt{\pi})^{-1/2} H_n(y) e^{-(1/2)y^2}. \quad (\text{A } 2)$$

The set $\{\phi_n\}$ is orthonormal, in the sense that

$$\langle \phi_i, \phi_j \rangle = \delta_{ij} \quad (\text{A } 3)$$

where δ_{ij} is the Kronecker delta. The set also forms a basis for square-integrable functions $F(y)$ over $(-\infty, \infty)$, i.e.

$$F(y) = \sum_{n=0}^{\infty} F_n \phi_n, \quad \text{with } F_n = \langle F, \phi_n \rangle \quad (\text{A } 4)$$

Two useful recursion relations are

$$\phi'_n = \sqrt{\frac{n}{2}} \phi_{n-1} - \sqrt{\frac{n+1}{2}} \phi_{n+1} \quad (\text{A } 5a)$$

and

$$\phi_{n+1} = \sqrt{\frac{2}{n+1}} y \phi_n - \sqrt{\frac{n}{n+1}} \phi_{n-1}. \quad (\text{A } 5b)$$

Combining the two gives

$$\phi'_n = -y \phi_n + \sqrt{2n} \phi_{n-1} \quad (\text{A } 6a)$$

$$\phi'_n = y \phi_n - \sqrt{2(n+1)} \phi_{n+1}. \quad (\text{A } 6b)$$

A.1. Inversion of \mathcal{L}_2

Consider a differential equation of the form

$$\mathcal{L}_2 \phi = F(y), \quad (\text{A } 7)$$

with $F(y)$ again being a square integrable function over $(-\infty, \infty)$. The solution ϕ can be determined uniquely and expressed in terms of an infinite series:

$$\phi = \sum_{n=0}^{\infty} a_n \phi_n \quad \text{with } a_n = -\frac{1}{2n+1} \langle F, \phi_n \rangle. \quad (\text{A } 8)$$

We can see that the above is equivalent to

$$\phi = \int_{-\infty}^{\infty} F(s) \cdot \sum_{n=0}^{\infty} \frac{-1}{2n+1} \phi(s) \phi(y) ds. \quad (\text{A } 9)$$

Defining the Green's function as $G(s; y) = -\phi(s)\phi(y)/(2n+1)$, the inverse operator \mathcal{L}_2^{-1} can be formally defined as an integral operator:

$$\mathcal{L}_2 \phi = F \iff \phi = \mathcal{L}_2^{-1} F = \int_{-\infty}^{\infty} F(s) G(s; y) ds. \quad (\text{A } 10)$$

Appendix B. Derivation of (5.13)

With the weakly nonlinear scaling,

$$\mathcal{L} = -\frac{\partial^2}{\partial y^2} + \frac{y(y + \epsilon u_{0y})}{1 + \epsilon \eta} = -\mathcal{L}_2 - \epsilon y(-u_{0y} + y\eta) + O(\epsilon^2). \quad (\text{B } 1)$$

Since $u_{0y} = -\eta_y/y$, we can replace $-u_{0y} + y\eta$ by $\mathcal{L}_1\eta$. Solving $\mathcal{L}\phi = F$ asymptotically by expanding $\phi = \phi_0 + \epsilon\phi_1 + \dots$, the first two terms are given by

$$\phi_0 = -\mathcal{L}_2^{-1}F \quad (\text{B } 2a)$$

$$\phi_1 = \mathcal{L}_2 y(u_{0y} - y\eta)\phi_0. \quad (\text{B } 2b)$$

We thus can write

$$\mathcal{L}^{-1} = -\mathcal{L}_2^{-1} - \epsilon (\mathcal{L}_2^{-1}y(u_{0y} - y\eta)) \mathcal{L}_2^{-1} + O(\epsilon^2). \quad (\text{B } 3)$$

Appendix C. Symmetric elliptic differential operators

Here we provide a brief overview of symmetric elliptic differential operators. For details see Evans (2010). The differential operator of the form

$$\mathcal{L} = -\sum_{i,j=1}^2 \frac{\partial}{\partial x_i} \left(a_{ij} \frac{\partial}{\partial x_j} \right) \quad (\text{C } 1)$$

with variable coefficients a_{ij} , is symmetric if $a_{ij} = a_{ji}$. A symmetric operator is uniformly elliptic if the coefficient matrix A , with elements a_{ij} , is positive-definite with eigenvalues greater than or equal to a constant $K > 0$.

For a symmetric matrix A to be positive-definite, it is sufficient to have $a_{11} > 0$ and $a_{11}a_{22} - a_{12}^2 > 0$. The smaller of the two real eigenvalues can be written as

$$\begin{aligned} \lambda &= \frac{1}{2} \left(a_{11} + a_{22} - \sqrt{(a_{11} + a_{22})^2 - 4(a_{11}a_{22} - a_{12}^2)} \right) \\ &= \frac{2(a_{11}a_{22} - a_{12}^2)}{a_{11} + a_{22} + \sqrt{(a_{11} + a_{22})^2 - 4(a_{11}a_{22} - a_{12}^2)}} \\ &\geq \frac{a_{11}a_{22}}{a_{11} + a_{22}}. \end{aligned} \quad (\text{C } 2)$$

When $a_{11}a_{22} - a_{12}^2 \geq K$ for some positive constant K , we must have $a_{11}a_{22} \geq K$. It is then possible to show that the last expression in (C2) attains a minimum value of $\sqrt{K}/2$, and thus λ is bounded from below by a positive content. We can therefore conclude that

$$a_{11}a_{22} - a_{12}^2 \geq K > 0 \quad (\text{C } 3)$$

is a sufficient condition for a symmetric operator to be uniformly elliptic.

Uniformly elliptic operators with the form of (C1) have similar properties as Sturm–Liouville operators: they have real eigenvalues which form an increasing sequence, with the smallest eigenvalue $\lambda_1 > 0$. In other words, the homogeneous problem $\mathcal{L}\phi = 0$ only has the trivial solution. The Fredholm alternative, which applies to uniformly elliptic operators, then guarantees that $\mathcal{L}\phi = F$ has a unique solution. We can therefore write $\phi = \mathcal{L}^{-1}F$. In addition the eigenfunctions ϕ_n of \mathcal{L} form an orthonormal basis that spans the space of square-integrable functions, allowing us to express the solution as

$$\phi = \sum_{n=0}^{\infty} a_n \phi_n \quad \text{with } a_n = \frac{1}{\lambda_n} \int_{\mathcal{G}} F \phi_n \, dx_1 \, dx_2. \quad (\text{C } 4)$$

REFERENCES

- ANDREWS, D. G., HOLTON, J. R. & LEOVY, C. B. 1987 *Middle Atmosphere Dynamics*. Academic.
- BELL, M. J., MARTIN, M. J. & NICHOLS, N. K. 2004 Assimilation of data into an ocean model with systematic errors near the equator. *Q. J. R. Meteorol. Soc.* **130**, 873–893.
- BOLIN, B. 1955 Numerical forecasting with the barotropic model. *Tellus* **7**, 27–49.
- BURGERS, G., BALMASEDA, M. A., VOSSEPOEL, F. C., VAN OLDENBORGH, G. J. & VAN LEEUWEN, P. J. 2002 Balanced ocean-data assimilation near the equator. *J. Phys. Oceanogr.* **32**, 2509–2519.
- CHARNEY, J. G. 1955 The use of the primitive equations of motion in numerical prediction. *Tellus* **7**, 22–26.
- CHARNEY, J. G. 1963 A note on large-scale motions in the tropics. *J. Atmos. Sci.* **20**, 607–608.
- DALEY, R. 1993 *Atmospheric Data Analysis*. Cambridge University Press.
- DERBER, J. & BOUTTIER, F. 1999 A reformulation of the background error covariance in the ECMWF global data assimilation system. *Tellus A* **51** (2), 195–221.
- EVANS, L. C. 2010 *Partial Differential Equations*. American Mathematical Society.
- FISHER, M. 2003 Background error covariance modelling. In *Proceedings of ECMWF Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean*, pp. 45–63.
- FUJIWARA, M., IWASAKI, S., SHIMIZU, A., SHIOTANI, M., MATSUURA, H., INAI, Y., HASEBE, F., MATSUI, I., SUGIMOTO, N., OKAMOTO, H., YONEYAMA, K., HAMADA, A., NISHI, N. & IMMLER, F. 2009 Cirrus observations in the tropical tropopause layer over the Western Pacific. *J. Geophys. Res.* **114**, D09304.
- GENT, P. R. & MCWILLIAMS, J. C. 1983 The equatorial waves of balanced models. *J. Phys. Oceanogr.* **13**, 1179–1192.
- GILL, A. E. 1980 Some simple solutions for heat-induced tropical circulation. *Q. J. R. Meteorol. Soc.* **106**, 447–462.
- GILL, A. E. 1982 *Atmosphere–Ocean Dynamics*. Academic.
- HECKLEY, W. A. & GILL, A. E. 1984 Some simple analytical solutions to the problem of forced equatorial long waves. *Q. J. R. Meteorol. Soc.* **110**, 203–217.
- KISTLER, R., KALNAY, E., COLLINS, W., SAHA, S., WHITE, G., WOOLLEN, J., CHELLIAH, M., EBISUZAKI, W., KANAMITSU, M. & KOUSKY, V. *et al.* 2001 The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation. *Bull. Am. Meteorol. Soc.* **82**, 247–268.
- LEVITAN, B. M. 1988 Sturm–Liouville operators on the whole line, with the same discrete spectrum. *Math. USSR Sb.* **60**, 77–106.
- MAJDA, A. J. 2003 *Introduction to PDEs and Waves for the Atmosphere and Ocean*. American Mathematical Society.
- MAJDA, A. J. & KLEIN, R. 2003 Systematic multiscale models for the tropics. *J. Atmos. Sci.* **60**, 393–408.
- MATSUNO, T. 1966 Quasi-geostrophic motions in the equatorial area. *J. Met. Soc. Japan* **44**, 25–43.
- MOURA, A. D. 1976 Eigensolutions of linearized balance equations over a sphere. *J. Atmos. Sci.* **33**, 877–907.
- PARRISH, D. F. & DERBER, J. C. 1992 The national-meteorological-centres spectral statistical-interpolation analysis system. *Mon. Weath. Rev.* **120**, 1747–1763.
- RAYMOND, D. J. 1994 Nonlinear balance on an equatorial beta-plane. *Q. J. R. Meteorol. Soc.* **120**, 215–219.
- RIPA, P. 1983 General stability conditions for zonal flows in a one-layer model on the beta-plane or the sphere. *J. Fluid Mech.* **126**, 463–489.
- ROUNDY, P. E. & FRANK, W. M. 2004 A climatology of waves in the equatorial region. *J. Atmos. Sci.* **61**, 2105–2132.
- SAUJANI, S. 2005 Towards a unified theory of balanced dynamics. PhD thesis, University of Toronto.
- SCHUBERT, W. H., SILVERS, L. G., MASARIK, M. T. & GONZALEZ, A. O. 2009 A filtered model of tropical wave motions. *J. Adv. Model. Earth Syst.* **1**, 3.

- SPARC INTERNATIONAL PROJECT OFFICE, 2010 *SPARC Chemistry-Climate Model Validation*. (ed. V. Eyring, T. G. Shepherd, and D. W. Waugh), *SPARC Report No. 5* (WCRP No. 132, WMO/TD No. 1526), <http://www.sparc-climate.org/publications/sparc-reports/sparc-report-no5/>.
- STEVENS, D. E., KUO, H.-C., SCHUBERT, W. H & CIESIELSKI, P. E. 1990 Quasi-balanced dynamics in the tropics. *J. Atmos. Sci.* **47**, 2262–2273.
- TEMPERTON, C. & WILLIAMSON, D. L. 1981 Normal mode initialization for a multilevel grid-point model. Part I: linear aspects. *Mon. Weath. Rev.* **109**, 729–743.
- THEISS, J. & MOHEBALHOJEH, A. R. 2009 The equatorial counterpart of the quasi-geostrophic model. *J. Fluid Mech.* **637**, 327–356.
- TITCHMARSH, E. C. 1962 *Eigenfunction Expansions Associated with Second-order Differential Equations. Part I*. Clarendon.
- VALLIS, G. K. 2006 *Atmospheric and Oceanic Fluid Dynamics*. Cambridge University Press.
- VERKLEY, W. T. M. & VAN DER VELDE, I. R. 2010 Balanced dynamics in the tropics. *Q. J. R. Meteorol. Soc.* **136**, 41–49.
- 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.
- WEIDEMAN, J. A. & REDDY, S. C. 2000 A Matlab differentiation matrix suite. *ACM Trans. Math. Softw.* **26**, 465–519.
- WHEELER, M. & KILADIS, G. N. 1999 Convectively coupled equatorial waves: analysis of clouds and temperature in the wavenumber-frequency domain. *J. Atmos. Sci.* **56**, 374–399.
- WILLIAMSON, D. L. & TEMPERTON, C. 1981 Normal mode initialization for a multilevel grid-point model. Part II: nonlinear aspects. *Mon. Weath. Rev.* **109**, 744–757.