Diabatic balance model for the equatorial atmosphere

Article

Published Version

Diabatic Balance Model for the Equatorial Atmosphere

IAN H. CHAN

Department of Physics, University of Toronto, Toronto, Ontario, Canada

THEODORE G. SHEPHERD

Department of Meteorology, University of Reading, Reading, Berkshire, United Kingdom

(Manuscript received 25 July 2013, in final form 4 November 2013)

ABSTRACT

Using an asymptotic expansion, a balance model is derived for the shallow-water equations (SWE) on the equatorial β plane that is valid for planetary-scale equatorial dynamics and includes Kelvin waves. In contrast to many theories of tropical dynamics, neither a strict balance between diabatic heating and vertical motion nor a small Froude number is required. Instead, the expansion is based on the smallness of the ratio of meridional to zonal length scales, which can also be interpreted as a separation in time scale. The leading-order model is characterized by a semigeostrophic balance between the zonal wind and meridional pressure gradient, while the meridional wind v vanishes; the model is thus asymptotically nondivergent, and the nonzero correction to v can be found at the next order. Importantly for applications, the diagnostic balance relations are linear for winds when inferring the wind field from mass observations and the winds can be diagnosed without direct observations of diabatic heating. The accuracy of the model is investigated through a set of numerical examples. These examples show that the diagnostic balance relations can remain valid even when the dynamics do not, and the balance dynamics can capture the slow behavior of a rapidly varying solution.

1. Introduction

A popular view of tropical dynamics is the equatorial wave theory of Matsuno (1966), where the dynamics is described by the normal modes of the linearized primitive equations on the equatorial β plane. In addition to the inertia–gravity and Rossby wave modes that are also found in midlatitudes, the equatorial region also admits mixed Rossby–gravity (MRG) waves and Kelvin waves. A fundamental feature of these special equatorial waves is that they are equatorially trapped with a meridional extent characterized by the equatorial Rossby radius of deformation $L_R = \sqrt{c/\beta}$, where c is the gravity wave speed and $\beta \approx 2.3 \times 10^{-11}\,\mathrm{m^{-1}\,s^{-1}}$ is the gradient of the Coriolis parameter at the equator. Gill (1980) examined the tropical circulation induced by a steady heat source near the equator, and found that the response can be interpreted in terms of Kelvin and Rossby waves (see also Webster 1972).

The equatorial wave theory is currently used as the theoretical basis for a host of phenomena in the tropics, such as the El Niño–Southern Oscillation, the quasi-biennial oscillation, and the Madden–Julian oscillation (MJO; e.g., Baldwin et al. 2001; Kiladis et al. 2009). This view is further corroborated by many observational studies where structures resembling equatorial waves are found; for example, Wallace and Kousky (1968) identified planetary-scale Kelvin waves in the equatorial lower stratosphere from radiosonde data. More recently, spectral analysis of proxies for deep tropical convection (e.g., outgoing longwave radiation and precipitable water; Wheeler and Kiladis 1999; Roundy and Frank 2004), provides ample evidence of convectively coupled equatorial waves. In addition to convection, equatorial waves can also be excited via lateral coupling with extratropical dynamics (Zhang and Webster 1992; Zhang 1993; Hoskins and Yang 2000).

An alternative view of equatorial dynamics is one that is based on the scale analysis of Charney (1963), who concluded that synoptic-scale motions in the tropics are quasi horizontal and quasi nondivergent, in the sense that the horizontal divergence is typically much smaller

Corresponding author address: Ian Chan, Department of Physics, University of Toronto, 60 St. George St., Toronto, ON M5S 1A7, Canada.
E-mail: ianchan@atmosp.physics.utoronto.ca

DOI: 10.1175/JAS-D-13-0224.1

than the vorticity, even more so than in the midlatitude quasigeostrophic theory. This balanced, quasi-nondivergent view of equatorial dynamics appears to be at odds with the traditional equatorial wave theory, where convection or coupling with the extratropics generates divergent flow in the form of equatorial waves. However in an analysis of data from the Tropical Ocean and Global Atmosphere Coupled Ocean–Atmosphere Response Experiment (TOGA COARE), vortical motions were indeed found to dominate over divergent motions even at scales up to 2000 km (Yano et al. 2009); this is particularly so during MJO events, where convection is active. For these reasons Yano and Bonazzola (2009) argued that the convectively coupled equatorial waves may be reinterpreted in the framework of balanced (quasi nondivergent) dynamics, and the balanced dynamics may have a larger regime of validity than the scalings suggest.

The best known diabatic balance model for the tropics is the weak temperature gradient (WTG) model of Sobel et al. (2001), which is based on the shallow-water equations (SWE). Like many theories of tropical dynamics (e.g., Held and Hoskins 1985), it is characterized by a diagnostic balance between the horizontal divergence and mass source–sink (the analog of diabatic heating), which is denoted by Q:

$$H(u_x + v_y) = Q, \tag{1}$$

where H is the fluid depth at rest. [In this and subsequent equations, the x and y (and t) subscripts represent derivatives with respect to x and y (and t), respectively (except in variables L_y and L_x)]. The above balance has also been referred to as the ''free-ride balance'' by Fraedrich and McBride (1989). This truncation of the continuity equation filters out fast inertia–gravity waves (IGW) while retaining the slow, advective, balanced dynamics. Although the scaling underlying the WTG model is formally valid only on the mesoscale (Sobel et al. 2001; Majda and Klein 2003), the model has been used to study synoptic- and planetary-scale dynamics (e.g., Polvani and Sobel 2002; Bretherton and Sobel 2003; Zhou and Sobel 2006), and has been shown to be in qualitative agreement with previous work on the Gill model (Gill 1980) and Hadley circulation (e.g., Hsu and Plumb 2000).

The WTG model can be considered to be a diabatic generalization of the Charney (1963) model. However a crucial point is that the scaling does not require the divergence to be small (although of course there is a quasi-nondivergent regime). Instead, the key assumption to obtain a balance model is the constraint in Eq. (1). Because of this, applying the WTG model to large-scale dynamics cannot be entirely satisfactory as it filters out Kelvin waves along with IGW: in the adiabatic case, the balance relation in Eq. (1) reduces to $u_x + v_y = 0$, which is not satisfied by Kelvin waves. As Kelvin waves are fast on the subsynoptic scale, they should indeed be filtered in a balance model for those scales; on the other hand, Kelvin waves play a central role in low-frequency variability at the planetary scale (Wheeler and Kiladis 1999). Moreover, as pointed out by Delayen and Yano (2009), quasi nondivergence does not necessarily preclude Kelvin waves.

In this paper, we show that it is possible to construct a balance model for tropical planetary-scale dynamics that encompasses the low-frequency equatorial wave theory while remaining quasi nondivergent. The term ''balance'' is used henceforth to more generally describe the slow component of a geophysical model, while ''balance model'' is the simplified model that retains only the slow dynamics [for a more thorough discussion see Warn et al. (1995)].

Chan and Shepherd (2013, hereafter CS13) considered the SWE on the equatorial β plane, and showed that a family of balance models can be derived based on the modified asymptotic method of Warn et al. (1995). Their approach is based on the anisotropy of planetary-scale flows, and thus uses the ratio of meridional to zonal length scales (i.e., $L_y/L_x = \epsilon \ll 1$) as the small parameter. When nonlinearity is weak, the leading-order model corresponds to the linear theory for equatorial long waves and retains equatorial Rossby and Kelvin waves. In addition, the model suggests that the divergence in the slow dynamics is smaller than the vorticity by a factor of ϵ. These results are in agreement with Delayen and Yano (2009), who suggested that the traditional longwave model can potentially be reinterpreted within such an asymptotically nondivergent framework; however with the anisotropic approach the smallness of the divergence is not used as a balance constraint, thus allowing Kelvin waves in the balance model.

Our work here builds on the results of CS13 by including a mass source into the shallow-water equations. In section 2, we derive the corresponding diabatic longwave balance model using the slaving method. In particular, we show that the slow time scale for the fully nonlinear regime considered by CS13 can be reinterpreted as an advective time scale. The dynamics of the longwave balance model are explored through the classical Gill problem in section 3, and in section 4 we compare the longwave balance model to the classical WTG model. In section 5 we further explore the longwave balance model by considering a case where a steady mass source generates an unsteady flow with periodic shedding of vortices.

2. Derivation of the diabatic longwave balance model

We start with the shallow-water equations on the equatorial β plane, with the inclusion of Rayleigh drag in the momentum equations and a source Q in the mass equation:

$$u_t + uu_x + vu_y - \beta yv + gh_x = -au, \tag{2a}$$

$$v_t + uv_x + vv_y + \beta yu + gh_y = -av, \tag{2b}$$

$$h_t + uh_x + vh_y + h(u_x + v_y) = Q, \tag{2c}$$

where a is the damping coefficient. We adopt Rayleigh friction here to aid comparison with classical studies such as Gill (1980) and Sobel et al. (2001), and we note a can be set to 0 without affecting the subsequent derivation.

a. Nondimensionalization

We can nondimensionalize the system in the following manner:

$$x = L_x\tilde{x}, \quad y = L_y\tilde{y}, \quad t = \frac{1}{\epsilon\sqrt{\beta c}}\tilde{t},$$
$$(u, v) = U(\tilde{u}, \tilde{v}), \quad \text{and} \quad h = H\tilde{h}. \tag{3}$$

We expect the meridional scale L_y to be set by the equatorial Rossby radius of deformation $L_R = \sqrt{c/\beta}$. With a typical value of $c \approx 50\,\mathrm{m\,s^{-1}}$, we have $L_R = \sqrt{c/\beta} \approx 1500\,\mathrm{km}$ and the inertial period $1/\sqrt{\beta c}$ is approximately 8 h. The quantity ϵ is the anisotropy parameter defined as $\epsilon = L_y/L_x$, and we take $\epsilon \ll 1$ to be the key small parameter in our asymptotic expansion. For the first zonal mode, L_x is the circumference of the earth ($\approx$40 000 km), which results in $\epsilon \approx 0.038$ and a corresponding time scale of approximately 9 days. The parameter ϵ is related to the equatorial Rossby number $\mathrm{Ro} = U/\sqrt{\beta c}L_x$ and Froude number $\mathrm{Fr} = U/c$ via

$$\epsilon \equiv \frac{L_y}{L_x} = \frac{U}{\sqrt{\beta c}L_x}\frac{c}{U} = \frac{\mathrm{Ro}}{\mathrm{Fr}}. \tag{4}$$

We consider both $\mathrm{Fr} = 1$ and $\mathrm{Fr} \ll 1$ depending on the degree of nonlinearity. Though we set the Froude number in both cases, the former is in fact a more general scaling assumption as it makes less assumption regarding the size of various terms, and the $\mathrm{Fr} \ll 1$ limit is a regular rather than a singular approximation as it does not affect the order of the system. Our study is thus different from many other equatorial theories (such as Charney 1963) that rely heavily on a small Froude number to obtain the balance reduction, whereas in our case the key singular perturbation parameter is the anisotropy ϵ. With Eq. (4), the time scale can be rewritten as

$$\frac{1}{\epsilon\sqrt{\beta c}} = \frac{L_x}{c} = \frac{U}{c}\frac{L_x}{U} = \mathrm{Fr} \times \tau_{\mathrm{adv}}. \tag{5}$$

Thus, if $\mathrm{Fr} = 1$, the time scale is simply an advective time scale τ_{adv}. In this case Eq. (5) can alternatively be written as

$$\frac{1/\sqrt{\beta c}}{\tau_{\mathrm{adv}}} = \epsilon, \tag{6}$$

and since $1/\sqrt{\beta c}$ is the maximum time scale of inertia–gravity waves, this suggests that the anisotropy condition $\epsilon \ll 1$ can also be interpreted as a separation in time scale. We write the mass source as $Q = Q_0\tilde{Q}$ and Rayleigh damping as $a = \tilde{a}c/L_x$. After dropping the tildes, we have

$$u_t + \mathrm{Fr}\left(uu_x + \frac{1}{\epsilon}vu_y\right) - \frac{1}{\epsilon}yv + \frac{1}{\mathrm{Fr}}h_x = -au, \tag{7a}$$

$$v_t + \mathrm{Fr}\left(uv_x + \frac{1}{\epsilon}vv_y\right) + \frac{1}{\epsilon}\left(yu + \frac{1}{\mathrm{Fr}}h_y\right) = -av, \quad \text{and} \tag{7b}$$

$$h_t + \mathrm{Fr}\left[(hu)_x + \frac{1}{\epsilon}(hv)_y\right] = \mathrm{Fr}\frac{Q_0}{H}\frac{L_x}{U}Q. \tag{7c}$$

We will adopt a WTG-type scaling:

$$\frac{Q_0}{H} = \frac{U}{L_x}, \tag{8}$$

which suggests a balance between the diabatic heating and divergence; however unlike the WTG model, we do not require this to be the dominant balance in the mass equation. If the nonlinearity is strong and $\mathrm{Fr} = 1$, Eq. (7) becomes

$$u_t + uu_x + \frac{1}{\epsilon}vu_y - \frac{1}{\epsilon}yv + h_x = -au, \tag{9a}$$

$$v_t + uv_x + \frac{1}{\epsilon}vv_y + \frac{1}{\epsilon}(yu + h_y) = -av, \quad \text{and} \tag{9b}$$

$$h_t + (hu)_x + \frac{1}{\epsilon}(hv)_y = Q. \tag{9c}$$

When the nonlinearity is weak and $\mathrm{Fr} = \epsilon \ll 1$, it is convenient to express h as the unperturbed depth plus a small perturbation η: $h = 1 + \epsilon\eta$, and Eq. (7) becomes

$$u_t + \epsilon u u_x + v u_y - \frac{1}{\epsilon} y v + \eta_x = -a u, \tag{10a}$$

$$v_t + \epsilon u v_x + v v_y + \frac{1}{\epsilon}(y u + \eta_y) = -a v, \quad \text{and} \tag{10b}$$

$$\eta_t + \epsilon(\eta u)_x + (\eta v)_y + u_x + \frac{1}{\epsilon} v_y = Q. \tag{10c}$$

FORCING FUNCTION

The mass source Q frequently used in the literature can be interpreted as a Newtonian relaxation of the layer thickness toward a prescribed equilibrium h_e over a time scale of τ. In its dimensional form,

$$Q = \frac{h_e - h}{\tau}, \tag{11}$$

which can also be interpreted as a prescribed mass source h_e/τ balanced by a sink h/τ. Using the nondimensionalization $h = H(1 + \mathrm{Fr} \times \tilde{\eta}) = H\tilde{h}$, $h_e = H(1 + \mathrm{Fr} \times \tilde{\eta}_e) = H\tilde{h}_e$, and $\tau = \tau_0 \tilde{\tau}$, Q becomes

$$\mathrm{Fr}\frac{H}{\tau_0}\frac{\tilde{\eta}_e - \tilde{\eta}}{\tilde{\tau}} \quad \text{or} \quad \frac{H}{\tau_0}\frac{\tilde{h}_e - \tilde{h}}{\tilde{\tau}} \tag{12}$$

depending on whether $\mathrm{Fr} \ll 1$ or $\mathrm{Fr} = 1$. Note in either case the scaling assumption

$$\frac{Q_0}{H} = \frac{U}{L_x} \Rightarrow \frac{L_x/c}{\tau_0} = 1, \tag{13}$$

that is, the relaxation time scale is comparable to the slow time scale. Setting $\tau_0 = L_x/c$, the nondimensionalized forcing is given by

$$\tilde{Q} = \frac{\tilde{\eta}_e - \tilde{\eta}}{\tilde{\tau}} \quad \text{or} \quad \frac{\tilde{h}_e - \tilde{h}}{\tilde{\tau}}. \tag{14}$$

b. Derivation of the balance model

At this point we focus on the fully nonlinear regime, but we should point out that the derivation under the assumption $\mathrm{Fr} = \epsilon \ll 1$ is almost identical. Dropping the tildes in the second expression in Eq. (14), and defining $b = 1/\tau$, Eq. (9) can be rewritten as

$$(y u + h_y)_t + y u u_x + (h u)_{xy} + \frac{1}{\epsilon}\left[\frac{\partial^2}{\partial y^2} - \frac{y(y - u_y)}{h}\right] h v + y h_x = -a y u + b h_{ey} - b h_y, \tag{15a}$$

$$v_t + u v_x + \frac{1}{\epsilon} v v_y + \frac{1}{\epsilon}(y u + h_y) = -a v, \quad \text{and} \tag{15b}$$

$$h_t + (h u)_x + \frac{1}{\epsilon}(h v)_y = b(h_e - h), \tag{15c}$$

where Eq. (15a) is obtained by combining Eqs. (9a) and (9c). We seek balance relations of the following form:

$$u = u(h; \epsilon) = u_0(h) + \epsilon u_1(h) + \epsilon^2 u_2(h) + \cdots \tag{16a}$$

$$v = v(h; \epsilon) = v_0(h) + \epsilon v_1(h) + \epsilon^2 v_2(h) + \cdots, \tag{16b}$$

which tacitly views the wind field as a function of the fluid depth h. Gathering the leading order terms in Eqs. (15a) and (15b), we have

$$\mathcal{L}(h v_0) \equiv \left[-\frac{\partial^2}{\partial y^2} + \frac{y(y - u_{0y})}{h}\right] h v_0 = 0, \tag{17a}$$

$$v_0 v_{0y} + y u_0 + h_y = 0, \tag{17b}$$

where $\mathcal{L}$ is a differential operator. Although the equations appear to be coupled they are in fact not; as whenever the flow is inertially stable [i.e., $y(y - u_{0y}) > 0$ everywhere], the only solution to the homogeneous differential equation in Eq. (17a) is $v_0 = 0$ (for details, see CS13); furthermore, the operator can be inverted, that is, $\mathcal{L}\phi = F \Rightarrow \phi = \mathcal{L}^{-1}F$ for an arbitrary function $F(y)$. The implication is that Eq. (17b) simplifies to

$$u_0 = -\frac{1}{y}\frac{\partial h}{\partial y}, \tag{18}$$

or in other words, the zonal wind is in geostrophic balance with the meridional pressure gradient. From Eq. (15c) we can recover the $O(1)$ prognostic equation:

$$h_t + (h u_0)_x + (h v_1)_y = b(h_e - h). \tag{19}$$

At this point we should emphasize that we do not assume a WTG balance in Eq. (1) in the mass equation. To close the $O(1)$ model we have to compute v_1 by going to the next order in the asymptotic expansion. The next order terms in Eqs. (15a) and (15b) give

$$\mathcal{L}(h v_1) = y h_x + y u_0 u_{0x} + (h u_0)_{xy} + a y u_0 - b(h_e - h)_y, \tag{20a}$$

$$u_1 = 0. \tag{20b}$$

At this point, $\mathcal{L}$ is known as u_0 is obtained from h through geostrophic balance and thus Eq. (20a) can be inverted explicitly, giving us

$$v_1 = \frac{1}{h}\mathcal{L}^{-1}[yh_x + yu_0u_{0x} + (hu_0)_{xy} - (a-b)h_y - bh_{ey}]. \tag{21}$$

Together with $u_1 = 0$, we now have the balance relations up to $O(\epsilon)$. By collecting the $O(\epsilon)$ terms in Eq. (15a), we can show that $v_2 = 0$, and thus Eq. (15c) becomes

$$h_t + (hu_0)_x + (hv_1)_y + O(\epsilon^2) = b(h_e - h); \tag{22}$$

in other words, Eq. (19) is in fact the $O(\epsilon)$ prognostic equation. To summarize, the $O(\epsilon)$ balance model is given by

$$u = u_0(h) = -\frac{1}{y}\frac{\partial h}{\partial y}, \tag{23a}$$

$$v = \epsilon v_1(h) = \frac{\epsilon}{h}\mathcal{L}^{-1}[yh_x + yu_0u_{0x} + (hu_0)_{xy} - (a-b)h_y - bh_{ey}], \quad \text{and} \tag{23b}$$

$$h_t + (hu_0)_x + (hv_1)_y = b(h_e - h). \tag{23c}$$

With an asymptotic approach it is in principle possible to go to the next order to obtain an $O(\epsilon^2)$ balance model: the $O(\epsilon)$ terms from Eq. (15b) give $yu_2 = -v_{1t} - u_0v_{1x} + v_1v_{1y} + av_1$, where v_{1t} is obtained by differentiating Eq. (21) with respect to time, and h_t is subsequently eliminated using Eq. (23c); this results in a proper balance relation for u_2, which provides an ageostrophic correction to u.

c. Relationship to the Gill model

We now show that the nonlinear longwave balance model in Eq. (23) reduces to Gill's model in the limit of weak forcing. In this case, we write $h_e = 1 + \hat{\epsilon}\eta_e$ and $h = 1 + \hat{\epsilon}\eta$ in anticipation of small height perturbations due to the weak forcing, where $\hat{\epsilon} \ll 1$ can be interpreted as a small Froude number assumption; similarly we expect the forced wind field to be weaker: $(u, v) \to \hat{\epsilon}(u, v)$. With these rescalings, Eq. (23a) becomes

$$u = -\frac{1}{y}\frac{\partial \eta}{\partial y}. \tag{24a}$$

For Eq. (23b), first notice that $\mathcal{L} \approx -\partial_{yy} + y^2 + O(\hat{\epsilon})$. Using asymptotic expansion, one can show that the inverse $\mathcal{L}^{-1}$ is approximately given by $(-\partial_{yy} + y^2)^{-1} + O(\hat{\epsilon})$ (cf. appendix B of CS13). Defining $\mathcal{L}_2 = -\partial_{yy} + y^2$, Eq. (23b) becomes

$$v = \epsilon v_1(\eta) = \epsilon\{\mathcal{L}_2^{-1}[y\eta_x + u_{0xy} - (a-b)\eta_y - b\eta_{ey}]\}, \tag{24b}$$

where only the leading order terms are retained. Similarly by ignoring $O(\hat{\epsilon})$ and smaller terms, the prognostic equation in Eq. (23c) becomes

$$\eta_t + u_{0x} + v_{1y} = b(\eta_e - \eta). \tag{24c}$$

It can be shown that the linear wave dynamics of Eq. (24) is equivalent to the time-dependent Gill model used in Heckley and Gill (1984), but the advantage of our approach is that the meridional wind can be diagnosed directly from η via Eq. (24b), without having to decompose the variables in terms of Hermite functions.

3. Balanced dynamics

In this section we explore the dynamics of the nonlinear longwave balance model in the context of time-varying forcing, as prior studies on the longwave approximation (e.g., Gill 1980; Heckley and Gill 1984; Gill and Philips 1986; Van Tuyl 1986) focused mainly on the response to a steady forcing. Through the use of numerical calculations, we explore how the response of the SWE varies with different forcing time scales, and determine how well the longwave balance model captures the response.

a. Numerical model

With ϵ set to 1, both the full equatorial SWE and the nonlinear longwave balance model are solved numerically with the spatial derivatives approximated using a sixth-order finite-difference scheme in the zonal direction, and a pseudospectral method with Hermite basis functions in the meridional direction. The computational domain is zonally periodic, and the solutions are expected to decay exponentially away from the equator, which is satisfied automatically by the Hermite basis functions. The time derivatives are discretized using second-order Adams–Bashforth, with the first step calculated via fourth-order Runge–Kutta. As nonlinearity can lead to a steepening of Kelvin waves and subsequently shock formation in the inviscid case (see, e.g., Boyd 1980; Ripa 1982), a small viscosity term [e.g., $\nu(u_{xx} + u_{yy})$] is added to all three equations to ensure smoothness of the solutions by arresting the development of shocks.

For WTG type studies, the typical value chosen for the gravity wave speed is $c \approx 50\,\mathrm{m\,s^{-1}}$, which corresponds to a Rossby radius of deformation of $L_R = \sqrt{c/\beta} \approx 1500\,\mathrm{km}$ and an inertial period of $1/\sqrt{\beta c} \approx 8\,\mathrm{h^{-1}}$. The computational domain is $[-13.3, 13.3]$ in the zonal direction, which corresponds roughly to the circumference of the earth. Following Zhou and Sobel (2006), we choose $a = 0.1$ and $b = 1/8.64$, respectively, which approximately

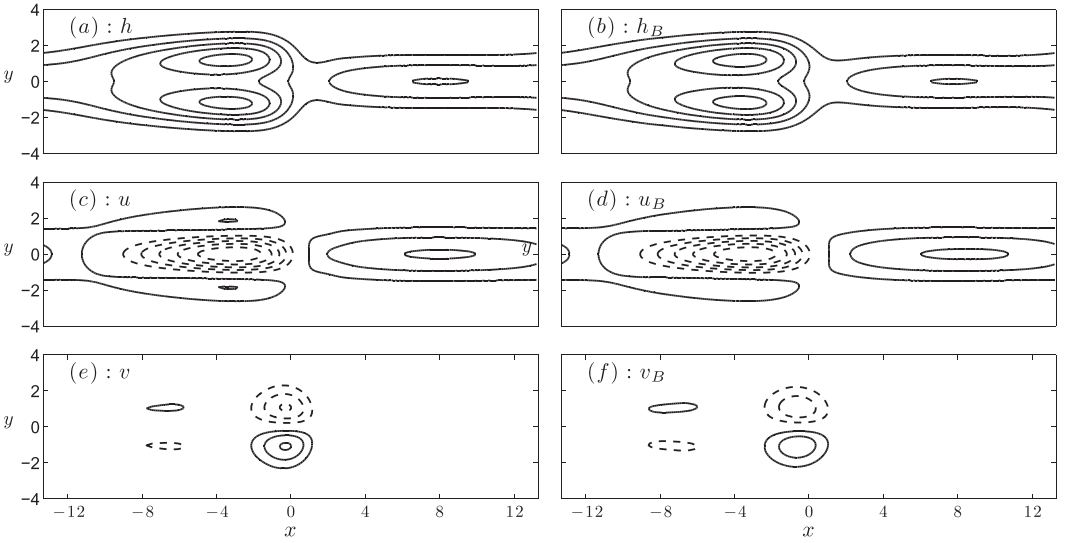

FIG. 1. A comparison between the responses of (a),(c),(e) the full model (no subscript) and (b),(d),(f) the balance model (subscript B) to a periodic Gaussian mass source located at the origin with a forcing period of $T = 90$. (a),(b) Perturbation in height field h; (c),(d) zonal wind u; and (e),(f) meridional wind v. Contour intervals are 0.01 for h and u and 0.005 for v. The zero contours are not shown for clarity.

correspond to a Rayleigh damping time scale of 3.5 days and a Newtonian relaxation time scale of 3 days.

Here we adopt the usual Newtonian relaxation to equilibrium as the model for a mass source. The equilibrium is chosen to take on a Gaussian shape centered at the origin:

$$h_e(x, y, t) = 1 + F(t) \exp[-(x^2 + y^2)/2], \qquad (25)$$

where the amplitude of the forcing $F(t)$ is allowed to vary in time.

b. Comparing full model and balanced model

We first choose $F(t) = \sin(2\pi t/90)$, where the period of 90 units corresponds to approximately 30 days; the models are integrated forward in time until a periodic solution emerges. In Fig. 1 the fields are plotted for $t = 135$, which is a point when the forcing $F(t) = 0$. The fields for the full model and balance model are plotted in the left and right columns, respectively, and they show excellent agreement in general.

The perturbation to the height field $\eta = h - 1$ for the full model and the balance model (denoted by subscript B) are plotted in Figs. 1a and 1b, respectively. We can see the double gyre associated with a Rossby wave to the west of the origin and a Kelvin wave to the east, but unlike the classical Gill problem, these waves propagate away from the forcing region because of the forcing being periodic in time.

We would like to determine the relative importance between the divergent and vortical motions. To aid comparison, we follow Delayen and Yano (2009) and define a root-mean-square (RMS) ratio:

$$r = \sqrt{\langle \delta^2 \rangle / \langle \zeta^2 \rangle} \quad \text{with} \quad \langle f^2(x, y, t) \rangle = \iiint f^2 \, dx \, dy \, dt, \qquad (26)$$

where the integration is carried out over the entire domain and over a period of forcing. For $T = 90$, we have $r = 0.22$, suggesting that the vorticity is significantly larger than the divergence. This ratio is larger than what is anticipated from linear wave theory (Delayen and Yano 2009), and comparable to the values seen in the TOGA COARE dataset (Yano et al. 2009).

We repeat the same calculation but with a forcing period reduced to $T = 10$, which corresponds to about 3 days; the results are plotted in Fig. 2. In the full model (left column) the primary response is in the form of Kelvin waves propagating to the east, while the Rossby wave response seen in steady-state models (e.g., Gill 1980) is absent. The balance model reproduces Kelvin wave dynamics to the east of the forcing region, which should not be surprising as in the linear limit the balance model Kelvin waves are represented exactly for all wavelengths (see CS13).

On the other hand, the balance model has a Rossby wave response to the west of the heat source that is not observed in the full SWE model. This spurious Rossby wave generation is due to the fact that the rollover of Rossby wave frequency at higher wavenumbers is not modeled correctly in the leading-order longwave balance

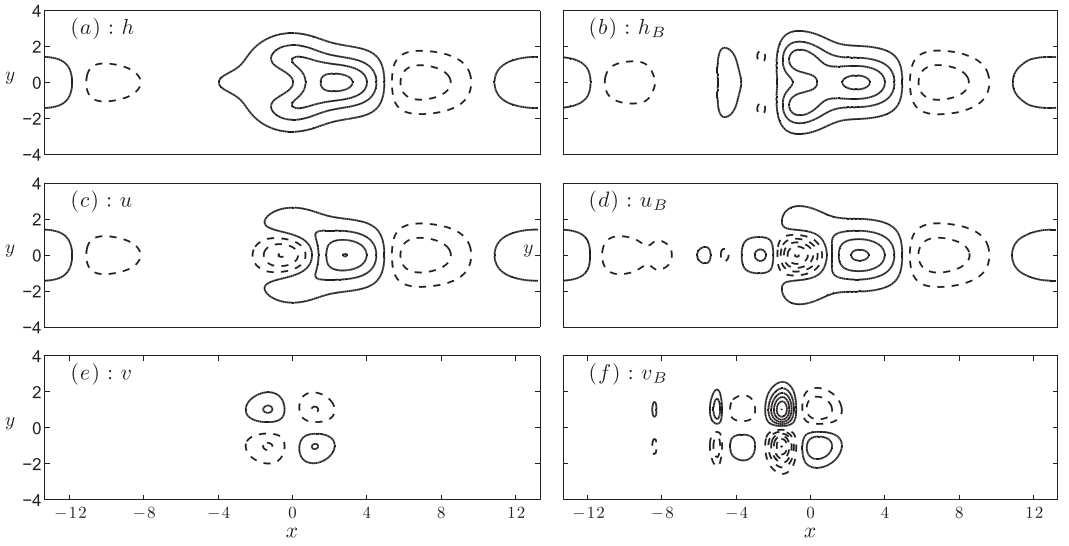

FIG. 2. As in Fig. 1, but for a forcing period of $T = 10$. Contour intervals are 0.02 for h and u and 0.01 for v.

model; instead, the frequency of Rossby waves increases linearly with wavenumber (see Stevens et al. 1990 and CS13), and the result is that a high-frequency forcing excites spurious Rossby waves.

It should be noted that for $T = 10$, the divergent flow becomes almost comparable to rotational flow, with the RMS ratio being $r = 0.81$. This should be expected as when the forcing period is reduced, the frequency of the Rossby and Kelvin waves excited will increase, and as a result k will have to increase as dictated by the dispersion relation. As a result, the generated waves become increasingly isotropic as the forcing period is reduced and the balance model diverges from the full SWE.

c. Balance relations

In the previous section we have shown that the balance model fails to track the full system for a small T, but as Warn et al. (1995) pointed out, it is possible for the balance relations to remain an accurate description of the slow dynamics even in the case where the balance dynamics fails. To determine if this is the case, we take the mass field from the full SWE runs, and use the balance relations in Eqs. (23a) and (23b) to calculate an inverted wind field $\mathbf{u}_I = (u_I, v_I)$; by comparing the inverted wind field to the actual wind field, we can determine whether the balance relations give a reasonable characterization of the full model.

For $T = 90$, u_I and v_I are plotted in Figs. 3a,b, respectively, and it is not surprising that the inverted wind field $\mathbf{u}_I$ is very similar to the wind field $\mathbf{u}$ in the full model (Figs. 1c,e) as the longwave balance model was able to reproduce the dynamics well. More surprising is that when the same calculation is repeated for $T = 10$, we find that $\mathbf{u}_I$ (Figs. 3c,d) is again very similar to $\mathbf{u}$ (Figs. 2c,e). The difference between the balance model and the wind field inverted via the balance relations is particularly

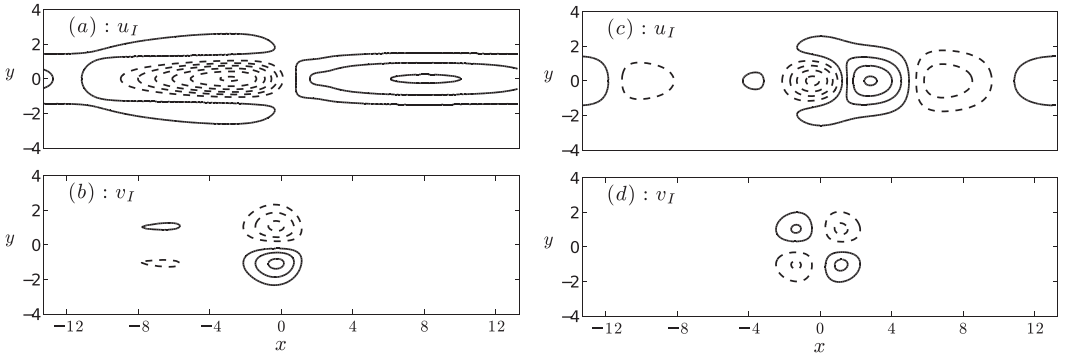

FIG. 3. Balance wind field $\mathbf{u}_I$ (u_I, v_I) inverted from mass field for (a),(b) $T = 90$ and (c),(d) $T = 10$ with (top) u_I and (bottom) v_I. Contour intervals are as in Figs. 1 and 2.

striking for the meridional wind: v_B (Fig. 2f) disagrees with v (Fig. 2e) to the west of the forcing region due to the spurious Rossby wave activity, whereas the inverted meridional wind v_I (Fig. 3d) is almost identical to v. The conclusion is that in the case of $T = 10$, the dynamics in the full SWE largely satisfy the balance relations in Eqs. (23a) and (23b) and is in fact close to the slow manifold defined by these balance relations.

4. Comparison with the Charney balance/weak temperature gradient model

Another view of balance dynamics in the tropics is exemplified by the theory of Charney (1963), which was subsequently generalized by Sobel et al. (2001) to result in the WTG model, and extended to the equatorial β plane by Zhou and Sobel (2006). While the WTG model is a balance model (as pointed out in the introduction), it filters out Kelvin waves, which our balance model retains. Apart from this important point, there are also significant differences, mainly in how the models are used from a diagnostic point of view. It is, therefore, useful to compare and contrast the two balance models in detail with the aid of some numerical examples.

The WTG model formally assumes a small Froude number, and is valid for length scales below the equatorial Rossby radius of deformation. Under these assumptions, the slow balanced dynamics is characterized by a balance between horizontal divergence and the mass source:

$$u_x + v_y = \delta = Q. \tag{27a}$$

The above equation can be used to replace δ_t in the divergence equation, resulting in a second diagnostic equation, which is the usual Charney balance generalized to the diabatic case:

$$\nabla^2\eta = -Q_t - \frac{1}{2}\nabla^2(\mathbf{u}\cdot\mathbf{u}) - [u(\zeta+y)]_y + [v(\zeta+y)]_x - a\delta. \tag{27b}$$

The balance dynamics is described by the vorticity equation:

$$\zeta_t + \mathbf{u}\cdot\nabla(\zeta+y) = -Q(\zeta+y) - a\zeta. \tag{27c}$$

In the adiabatic and inviscid regime, Eq. (27) simplifies to

$$u_x + v_y = 0, \tag{28a}$$

$$\nabla^2\eta = -(yu_\zeta)_y + (yv_\zeta)_x - \nabla(\mathbf{u}_\zeta\cdot\nabla\mathbf{u}_\zeta), \tag{28b}$$

$$\zeta_t + \mathbf{u}_\zeta\cdot\nabla(\zeta+y) = 0, \tag{28c}$$

where the subscript ζ denotes the vortical part of the wind field. Note that in this case the balance dynamics is nondivergent and satisfies the Charney balance in Eq. (28b).

The most significant difference between the models is that the simplification obtained in the nonlinear longwave balance model is due to anisotropy, which can be interpreted as a separation in time scale [cf. Eq. (6)], whereas the simplification obtained in the WTG model results from the assumption that the horizontal variation in temperature/mass is small, expressed in the smallness of Fr. An issue is that the use of Newtonian relaxation toward an equilibrium as the mass source [cf. Eq. (11)] is not formally consistent with the WTG scaling (Zhou and Sobel 2006). On the other hand, it is not an issue for the nonlinear longwave balance model considered here as we make no formal assumption regarding the scale of the height perturbation (i.e., we take Fr = 1).

a. Vorticity equation

To further compare the WTG model to the nonlinear longwave balance model, it is useful to recast our model in terms of divergence δ and vorticity ζ. With the longwave scaling, we have

$$\delta = \epsilon u_x + v_y = v_{0y} + \epsilon(u_{0x} + v_{1y}) + \epsilon^2(u_{1x} + v_{2y}) + \cdots \tag{29a}$$

$$\zeta = \epsilon v_x - u_y = -u_{0y} + \epsilon(v_{0x} - u_{1y}) + \epsilon^2(v_{1x} - u_{2y}) + \cdots. \tag{29b}$$

Since the leading-order balance yields $v_0 = 0$, the flow is nondivergent to the leading order irrespective of diabatic heating; this is to be contrasted with the WTG model, where leading-order divergence only vanishes in the adiabatic case. In other words, the WTG model formally assumes $\zeta \sim \delta \sim Q$ in the case of diabatic heating [cf. Eq. (5) of Sobel et al.(2001)] whereas the nonlinear longwave balance model implies that $\zeta \gg \delta \sim Q$; the latter appears to be more consistent with observations and reanalysis data for large-scale circulations (Sardeshmukh and Hoskins 1987; Yano and Bonazzola 2009; Yano et al. 2009).

It can be shown that the longwave balance model can be written as a vorticity equation. From Eq. (29b), we can see that ζ is approximated by $\zeta_0 = -u_{0y}$ at leading order. Now differentiating Eq. (23c) with respect to y, we have

$$h_{yt} + (hu_0)_{xy} + (hv_1)_{yy} = b(h_e - h)_y. \tag{30}$$

Rearranging Eq. (20a), we can see that

$$(hu_0)_{xy} + (hv_1)_{yy} = yv_1(y + \zeta_0) - yh_x - yu_0u_{0x} - ayu_0 + b(h_e - h)_y. \quad (31)$$

The above equation can then be substituted into Eq. (30), and after dividing through by y and differentiating with respect to y once more, we have

$$-(-h_y/y)_{yt} + v_{1y}(y + \zeta_0) + v_1(y + \zeta_0)_y - h_{xy} - (u_0u_{0x})_y = au_{0y}. \quad (32)$$

Recognizing that $u_0 = -h_y/y$, and thus $-h_{xy} - (u_0u_{0x})_y = (yu_0)_x - (u_0u_{0y})_x = [u_0(y + \zeta_0)]_x$, we can see that Eq. (32) is equivalent to

$$\zeta_{0t} + u_0(y + \zeta_0)_x + v_1(y + \zeta_0)_y = -\delta_1(y + \zeta_0) - a\zeta_0. \quad (33)$$

It should be emphasized that Eq. (33) is the full vorticity equation, and not a linearization about a basic state that is at rest or a statement of Sverdrup balance. Sardeshmukh and Hoskins (1985) and Held and Hoskins (1985) argued that while linear Sverdrup balance works for the lower troposphere, the vorticity balance in the upper troposphere is essentially nonlinear; specifically, the relative vorticity is comparable to the planetary vorticity at upper levels. This supports the choice of Fr = 1 in our scaling: since the scales for relative vorticity and planetary vorticity are U/L_y and βL_y, respectively, the ratio between the two terms is given by $u/\beta L_y^2 = U/c$, which is just the Froude number.

b. Overlapping regime

The nonlinear longwave balance model and the WTG model are in fact connected through the seasonal planetary equatorial weak temperature gradient (SPEWTG) model of Majda and Klein (2003), as the SPEWTG model emerges from the two former models as a distinguished limit, and thus can be regarded as an overlapping regime. Starting from the classical WTG model given by Eq. (27), we introduce anisotropy via $x \to x/\hat{\epsilon}$ and $v \to \hat{\epsilon}v$, where $\hat{\epsilon} \ll 1$ is a small parameter. To maintain the same advective time scale we rescale time via $t \to t/\hat{\epsilon}$, and in turn the mass source and Rayleigh friction have to be rescaled via $Q \to \hat{\epsilon}Q$ and $a \to \hat{\epsilon}a$ to maintain consistency for the relaxation time scales. With these rescalings and discarding terms of $O(\hat{\epsilon})$ or smaller, Eq. (27) becomes

$$u_x + v_y = Q, \quad (34a)$$

$$yu + \eta_y = 0, \quad (34b)$$

$$\zeta_{0t} + \mathbf{u} \cdot \nabla(\zeta_0 + y) = -Q(\zeta_0 + y) - a\zeta_0, \quad (34c)$$

where $\zeta_0 = -u_y$. A notable point is that the zonal geostrophic balance in Eq. (34b) emerges from the generalized Charney balance in Eq. (27b). The above equations are identical to the SPEWTG model given by Eqs. (2.24)–(2.26) of Majda and Klein (2003).

On the other hand, the SPEWTG model assumes the Froude number to be small (as opposed to Fr = 1 in the nonlinear longwave balance model) and the height perturbation to be of order Fr^2 (i.e., $h = 1 + \mathrm{Fr}^2\eta$). In addition, the meridional length scale chosen in the derivation of the SPEWTG model is assumed to be smaller than the equatorial Rossby radius of deformation L_R by a factor of $\mathrm{Fr}^{1/2}$. Thus, to obtain the SPEWTG model from the nonlinear longwave balance model, we assume $\hat{\epsilon} = \mathrm{Fr} \ll 1$ and rescale $h \to 1 + \hat{\epsilon}^2\eta$, $(u, v) \to \hat{\epsilon}(u, v)$, and $(x, y, t) \to \sqrt{\hat{\epsilon}}(x, y, t)$; for consistency we also rescale $a \to a/\sqrt{\hat{\epsilon}}$ and the mass source $Q = b(h_e - h)$ via $Q \to \sqrt{\hat{\epsilon}}Q$. Note that geostrophic balance for the zonal wind in Eq. (23a) remains invariant under this rescaling:

$$u_0 = -\frac{1}{y}\frac{\partial \eta}{\partial y}. \quad (35)$$

Expanding Eq. (20a) and rearranging, we have

$$u_{0xy} + v_{1yy} = Q_y - \hat{\epsilon}^2[(\eta v_1)_{yy} - y(y - u_{0y})v_1 + y\eta_x + yu_0u_{0x} + (\eta u_0)_{xy}] + \hat{\epsilon}ayu_0. \quad (36)$$

We can see that the $O(1)$ terms result in the usual WTG balance:

$$u_{0x} + v_{1y} = Q. \quad (37)$$

Finally, the rescaling leaves the vorticity equation in Eq. (33) invariant. We can see that Eqs. (33), (35), and (37) are identical to Eq. (34).

c. Balance relations as a diagnostic tool

We have established that the nonlinear longwave balance model and the classical WTG model are in fact connected in a special regime, but we also wish to highlight some important differences. A fundamental difference between the nonlinear longwave balance model and the WTG lies in the variable that is used to describe the evolution of the slow dynamics. In the WTG model, the slow dynamics is governed by vorticity, whereas in our model the fundamental variable is mass. Specifically, the mass/thermodynamic equation simplifies to a diagnostic equation under the WTG paradigm, whereas in the nonlinear longwave balance model it is used as the main prognostic equation.

A consequence is that there is a substantial difference in how the balance relations are used diagnostically. In

the WTG model, the horizontal divergence is first diagnosed from a given diabatic heat source, and subsequently used in conjunction with the vorticity (which is assumed to be known) to infer η via the generalized Charney balance relation in Eq. (27b). In practice, however, available observations are predominantly of the mass field (particularly in the tropics), and thus balance relations are more frequently used to infer the wind field based on mass observations, which results in a differential equation that is nonlinear in the winds. To further complicate the problem, diabatic heating is typically not observed, and thus the divergent wind field cannot be diagnosed. This then poses additional challenges for the use of Charney balance, as it is not possible to diagnose both the vortical and divergent part of the wind field using a single equation.

In contrast, in the construction of the nonlinear longwave balance model presented here, the wind field is tacitly assumed to be a function of the mass variable. Furthermore, the inversion of zonal and meridional winds are decoupled: we first use geostrophic balance in Eq. (23a) to obtain the zonal wind for a given mass field, which is then used in Eq. (23b) to obtain the meridional wind. A key feature of these balance relations is that they are both linear in the wind fields, and thus they are easier to apply than Charney balance. Finally, it is important to note that the diabatic heating is only needed for the $O(\epsilon)$ correction for the meridional wind, and u can be inverted to the leading order using mass observations alone; thus, these balance relations partially alleviate the problem of not having sufficiently good observational constraints on diabatic heating.

Furthermore, as modern data assimilation methods are all four-dimensional, h_t can be treated as an additional observed quantity. In this case, the meridional wind can be diagnosed via the zonal momentum equation in Eq. (10a):

$$v_1 = \frac{u_{0t} + u_0 u_{0x} + a u_0 - h_x}{y - u_{0y}}, \tag{38}$$

as u_0 and u_{0t} can be diagnosed separately from h and h_t, respectively, via geostrophic balance. Note that Eq. (38) can also be obtained by eliminating Q using Eqs. (10b) and (10c). Although Eq. (38) is not a proper balance relation as it contains a time derivative, it nonetheless allows us to circumvent the issue of incomplete information on diabatic heating and diagnose v. A similar approach is, however, not possible with the WTG model, as inferring the winds from h requires the use of Charney balance in Eq. (27b), where δ and ζ are coupled and cannot be determined independently of one another.

d. Numerical experiments

To further compare the WTG and nonlinear longwave balance model, we now compare the divergence computed using the WTG approximation in Eq. (27a) compared to the one computed using the nonlinear longwave balance model, as this is one of the main differences between the two models. Note that we follow Zhou and Sobel (2006) and allow for the use of Newtonian cooling in the WTG balance. The parameters used are identical to section 3a. Given a steady heating, we expect the WTG balance and nonlinear longwave balance model to behave similarly as the time derivative vanishes in the mass equation, and thus the dominant balance must reduce to the WTG balance unless the nonlinearity is strong. On the other hand, when the solution varies in time, the time derivative of the mass field may not be neglected as it is in the WTG theory.

1) STEADY MASS SOURCE

We first consider the standard Gill problem by setting $F(t) = 1$ in Eq. (32), and integrate the full SWE forward in time until a steady solution is obtained. In Figs. 4a–c, we plot the divergence δ obtained from the wind field of the full SWE model, the divergence δ_{WTG} estimated using the WTG approximation in Eq. (27a), and the divergence δ_I using the wind field inverted from the mass field via the nonlinear longwave balance relations in Eqs. (23a) and (23b). Note that both balance relations are able to reproduce the divergence well. To quantify the error, we introduce the relative square error (RSE):

$$\mathrm{RSE}(\delta_{\mathrm{WTG}}) = \frac{\iint (\delta_{\mathrm{WTG}} - \delta)^2 \, dx \, dy}{\iint \delta^2 \, dx \, dy}, \tag{39}$$

where the integral is carried out over the entire domain. For the WTG approximation we have $\mathrm{RSE}(\delta_{\mathrm{WTG}}) = 0.011$, while for the nonlinear longwave balance model $\mathrm{RSE}(\delta_I) = 0.005$. Note that for the steady solution h_t vanishes, and thus the difference between the two balance models lies in the nonlinear terms in the mass equation; however, in this case the maximum amplitude of η and u suggests that $\mathrm{Fr} \approx O(10^{-1})$, and thus it is not surprising that the two models agree as the nonlinearity is relatively weak.

We are also interested in determining how much of the divergence field can be reconstructed based on the mass field h alone, without additional information regarding the mass source $Q = b(h_e - h)$. For the WTG model this is impossible, as the divergence is entirely diagnosed via Q; on the other hand, u is completely

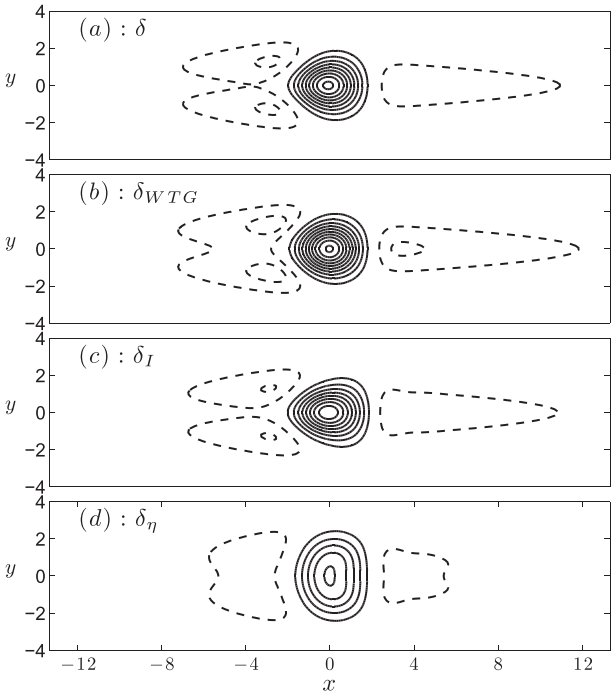

FIG. 4. A comparison of divergence between (a) the full SWE δ, (b) that obtained using the WTG approximation δ_{WTG}, and (c) that obtained via the nonlinear longwave balance model δ_I. Positive (negative) values are indicated by solid (dashed) contours with increments of 0.01 (-0.005). The zero contours are not shown for clarity. (d) Divergence δ_η reconstructed via the nonlinear longwave balance relations with height perturbation η while ignoring the mass source η_e.

determined through h in the nonlinear longwave balance model, while v only depends partially on Q. We thus calculate the divergence using Eqs. (23a) and (23b) without the terms $b(h_e - h)$, and the inverted divergence is plotted in Fig. 4d. The RSE in this case is 0.146, and the maximum divergence near the origin is about 55% of the actual value.

2) TIME-DEPENDENT SOLUTION

To investigate the effect of a nonsteady solution, we allow h_e to vary in time by choosing $F(t) = 1 + 0.5\sin(2\pi t/T)$ and integrate the model until a periodic solution emerges. In Fig. 5 we again plot the divergence δ, WTG divergence δ_{WTG}, and δ calculated from the nonlinear longwave balance relations for $T = 21$, which corresponds to a forcing period of roughly 7 days, and $t = 73$. We can see a significant disagreement between δ_{WTG} and δ, particularly away from the forcing region, and $\mathrm{RSE}(\delta_{\mathrm{WTG}}) = 0.26$; in contrast $\mathrm{RSE}(\delta_I) = 0.014$ at the same point in time. As the RSE varies significantly over a period of forcing, we also averaged RSE over a period of forcing and obtained $\overline{\mathrm{RSE}}(\delta_{\mathrm{WTG}}) = 0.14$ and $\overline{\mathrm{RSE}}(\delta_I) = 0.017$, suggesting that the nonlinear balance

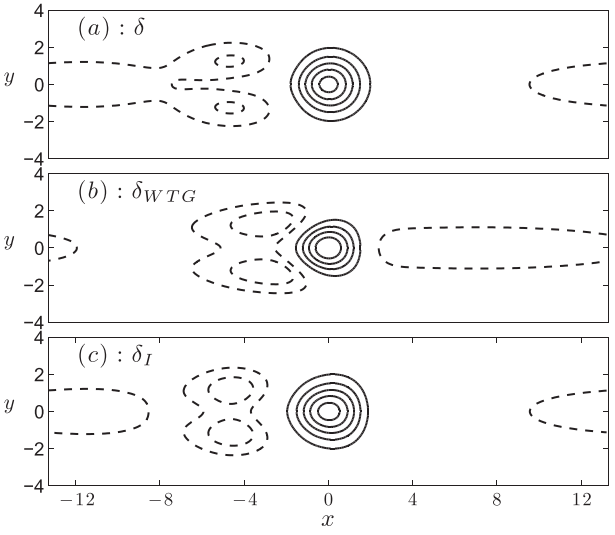

FIG. 5. As in Figs. 4a–c, but for a forcing with period $T = 21$. The divergence fields shown are for $t = 73$.

relations describe the divergence field more accurately compared with the WTG balance relation.

To further investigate how well the balance relations describe the dynamics of the full system, we varied the forcing period T from 10 to 240, which roughly corresponds to a time-scale range from 3 to 80 days, and computed $\overline{\mathrm{RSE}}$. The results, plotted in Fig. 6, show that the error for δ_I (squares) is smaller than δ_{WTG} (dots) over the range of forcing time scales tested. As T increases, $\overline{\mathrm{RSE}}$ converges to the value for the steady forcing (indicated by the dashed lines). It should be noted that the RSE for δ_{WTG} steadily increases as the forcing period decreases, whereas the RSE for δ_I increases much more slowly. As the amplitude of the height perturbation and wind field is similar in magnitude across all forcing periods, the difference between the two is presumably due to the inclusion of h_t in the nonlinear longwave balance model.

e. Hadley cell circulation

The shallow-water system in Eq. (9) with the mass source in Eq. (11) is also used as a model for an axisymmetric Hadley cell circulation (e.g., Polvani and Sobel 2002). A difference between the Gill problem and the Hadley cell problem is that the dynamics of the former is largely linear due to the relatively strong Rayleigh damping, whereas the latter problem is inherently nonlinear in the inviscid limit (i.e., Rayleigh friction vanishes). Hsu and Plumb (2000) considered the shallow-water analog of the classical Held–Hou model (Held and Hou 1980) on the f plane, while Polvani and Sobel (2002) examined a similar problem in Cartesian coordinates and on the equatorial β plane.

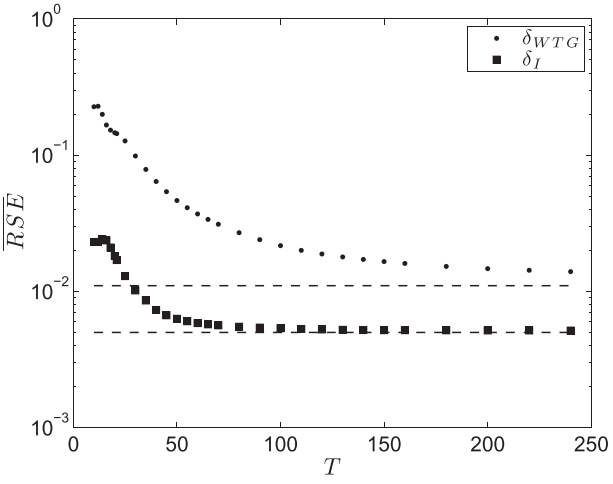

FIG. 6. Time-averaged relative square error $\overline{\text{RSE}}$ for δ_{WTG} and δ_I. The dashed lines are the RSEs for steady heating.

The WTG approximation has been applied to the Hadley cell problem by Polvani and Sobel (2002) and Zhou and Sobel (2006). In Zhou and Sobel (2006), the divergence is diagnosed from the mass source:

$$v_y = b(h_e - h), \tag{40}$$

where the full variation in h is used despite being formally inconsistent with the WTG scaling. When h is given, Eq. (40) can be integrated to directly infer the meridional wind. Zhou and Sobel (2006) solved the Hadley cell problem numerically with an axisymmetric mass source with a Gaussian profile centered at the equator:

$$h_e = 1 + \exp(-y^2/2) \tag{41}$$

for both the WTG and SWE systems. Although the WTG model qualitatively reproduces the steady-state solution to the SWE, the zonal and meridional winds are overestimated by 20% and 50%, respectively [cf. Fig. 1 of Zhou and Sobel (2006)].

Here we repeat the same calculation by integrating the nonlinear longwave balance model in Eq. (23) with Eq. (41) until it reaches a steady state. The results are plotted in Fig. 7 together with the solution from the full SWE, and they are indistinguishable from one another. In the right panel, we have also plotted the meridional wind field v_{WTG} that is inverted via the WTG balance in Eq. (40), and the meridional wind field inverted using the linearized balance relation in Eq. (24b); in both cases the balance relations significantly overestimate the meridional wind. This should not be surprising as the height perturbation η is about 0.6, and therefore the effects of nonlinearity will be significant but the nonlinear terms in the mass equation are ignored under the WTG approximation. Note that we have also computed the wind field $\mathbf{u}_I$ inverted using the nonlinear longwave balance relations and they are indistinguishable from $\mathbf{u}$ and $\mathbf{u}_B$ and hence omitted from Fig. 7 for clarity.

5. Eddy shedding

In the previous sections we have examined time-dependent solutions generated by a forcing that is periodic in time. It is also possible to generate an unsteady

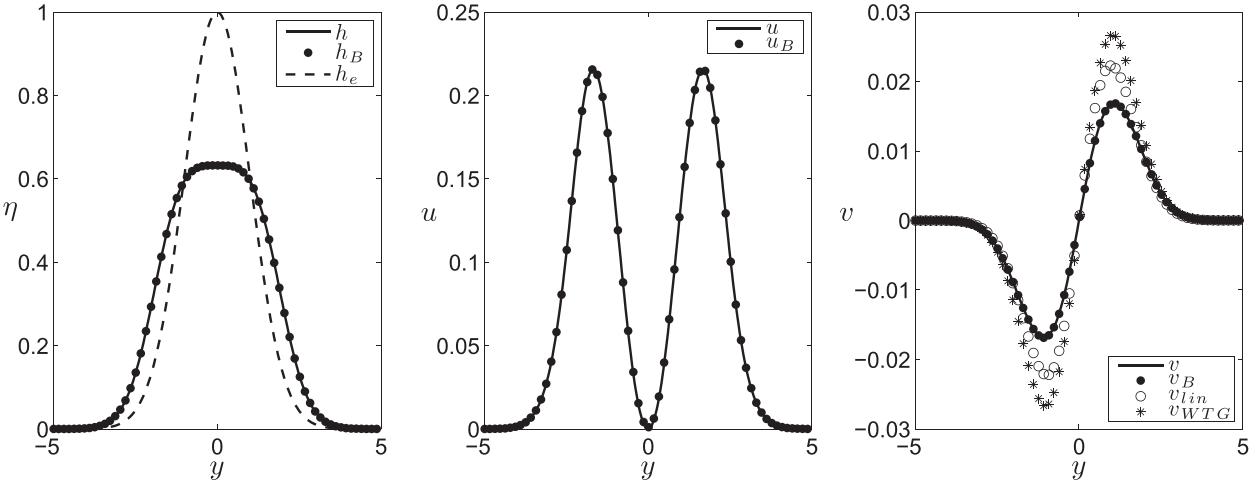

FIG. 7. Results from the Hadley circulation experiment: (left) η, (middle) u, and (right) v, each vs y. In all three panels, the solid line indicates the solution from the full SWE, while the black circles indicate the solution from the nonlinear longwave balance model. (left) The equilibrium height h_e is also plotted (dashed line). (right) The meridional velocity inverted from the linear (open circles) and WTG (asterisks) balance relations are also plotted.

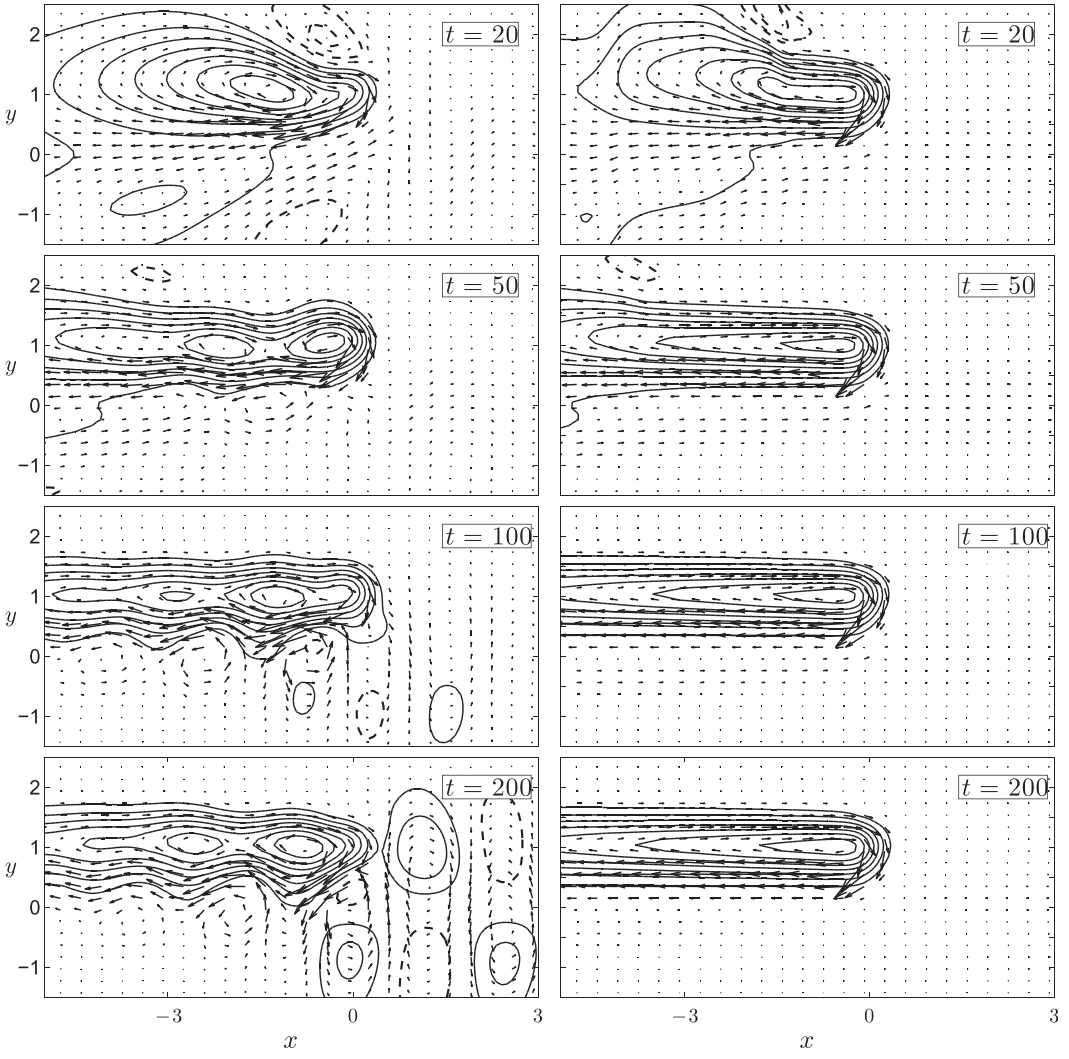

FIG. 8. (top to bottom) Snapshots from the vortex shedding experiment at $t = 20$, 50, 100, and 200. The results from the (left) full SWE and (right) nonlinear longwave balance model are plotted. The contours are the heights with a contour interval of 0.005. The positive (negative) values are indicated by solid (dashed) lines and the zero contour is not shown for clarity. The arrows indicate the wind field.

solution with a steady mass source: an example is vortex shedding examined by Hsu and Plumb (2000). Using a shallow-water model, Hsu and Plumb found that in the presence of a planetary vorticity gradient, the anticyclone resulting from a mass source becomes unstable and periodically sheds eddies. Zhou and Sobel (2006) repeated a similar set of calculations using the WTG model, and demonstrated that it was able to reproduce the eddy shedding well.

It is unclear whether this vortex shedding behavior can be reproduced in the longwave model, since the mass source used has a length scale significantly smaller than L_R, and thus is inconsistent with the scaling underlying the nonlinear longwave balance model. On the other hand, the anticyclone associated with the diabatic heating is elongated and is thus anisotropic and consistent with the scaling. We investigate this by comparing the nonlinear longwave balance model to the full SWE with a narrow Gaussian mass source situated off the equator:

$$h_e = 1 + Q_0 \exp\left\{-\frac{1}{2}\left[\left(\frac{x}{0.2}\right)^2 + \left(\frac{y - y_0}{0.2}\right)^2\right]\right\}. \quad (42)$$

For the numerical experiments, we choose $Q_0 = 1.5$, $a = 0.001$, $y_0 = 1$, and choose the viscosity $\nu = 4e^{-4}$ to match the parameters used in the previous studies.

The results from the full SWE are shown in the left column of Fig. 8. The mass source is switched on at $t = 0$ and an anticyclone develops to the west of the mass source

(top panel). The anticyclone elongates zonally and eventually splits into two distinct vortices; this can be seen clearly at $t = 50$ (second panel). The vortices propagate to the west and new vortices form continuously near the mass source. At later times ($t = 100$ and 200) MRG waves begin to appear to the east of the mass source, and they are characterized by a strong meridional cross-equatorial flow.

In the right column of Fig. 8 we present a similar experiment but with the nonlinear longwave balance model. Comparing the top two panels, we can see that the nonlinear longwave balance model behaves qualitatively similarly to the full SWE model; in particular, the strength of the anticyclonic circulation as well as the magnitude of the height perturbation are in good agreement. However, at later times the behavior of the two models clearly diverges, as neither the vortex shedding nor the MRG response is captured by the balance model; the latter is not surprising as the balance model filters out MRG waves. The balance model eventually converges to a steady solution, and it is clear that in this case the solution in the full SWE no longer follows the slow manifold as defined by the nonlinear longwave balance relations.

However, even in cases where the dynamics of the full system no longer converges to the slow manifold, the fast dynamics can still be seen as rapid oscillations around a ''guiding center'' defined by the slow manifold (Van Kampen 1985). To determine whether we can interpret the vortex shedding dynamics in this manner, we apply a time average to the variables to filter the fast motions. The height field averaged between $t = 250$ to 300 (denoted by $\overline{h}$) is plotted in the top panel of Fig. 9. We can immediately notice the similarity between $\overline{h}$ and the steady-state solution of the balance model (bottom right panel in Fig. 8). Using the nonlinear balance relations in Eqs. (23a) and (23b), we calculate the balance wind field $\overline{\mathbf{u}}_I$ using $\overline{h}$ and plot the results in the second panel of Fig. 9. Note that to avoid a singularity in Eq. (23a), $\overline{h}$ has to be adjusted to ensure that the derivative $\overline{h}_y$ vanishes as $y \to 0$, and here we follow the same adjustment scheme used by CS2013 (see section 4.1.1 of CS13), which corresponds to filtering of residual MRG waves in $\overline{h}$. Comparing the time-averaged wind field from the full model $\overline{\mathbf{u}}$ (third panel of Fig. 9) to $\overline{\mathbf{u}}_I$ reveals that they are indeed similar, indicating that the time-averaged wind field is balanced.

The last panel in Fig. 9 is a comparison between the zonal wind u from the full model at $t = 300$ and the inverted zonal wind u_I for $y = 0.44$, which is approximately the latitude with maximum easterly winds. It is clear that the balanced/geostrophic component of the time-averaged zonal circulation acts as a guiding center for the fast small-scale oscillations in the form of vortices to the west and MRG waves to the east.

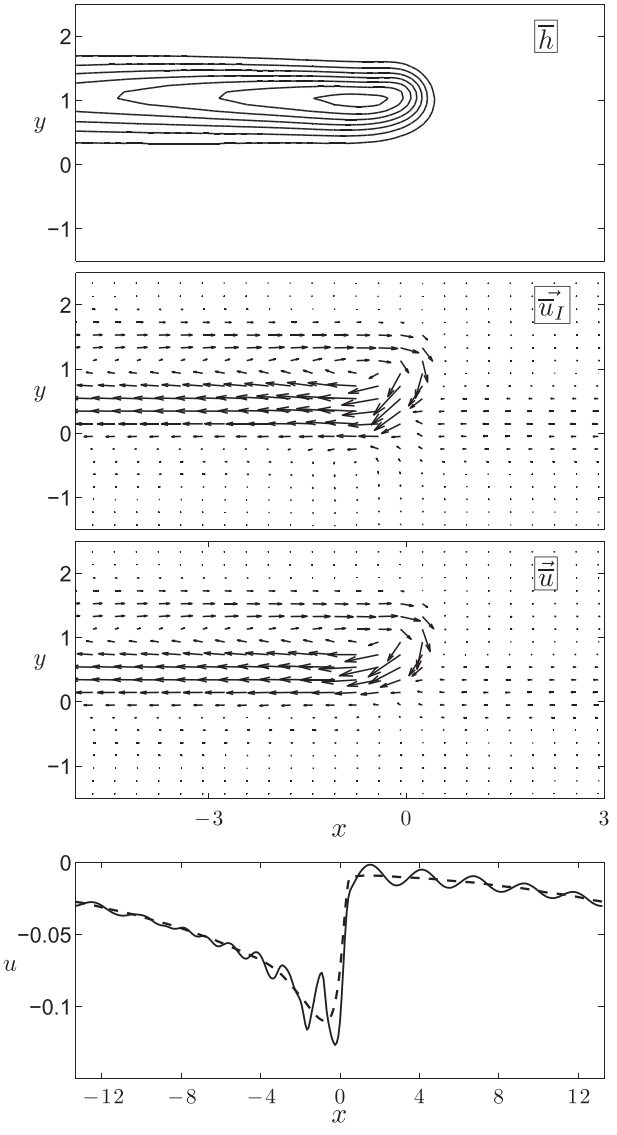

FIG. 9. (top to bottom) The time-averaged height $\overline{h}$ with a contour interval of 0.005 with positive (negative) values indicated by solid (dashed) lines and the zero contour not shown for clarity; the balance wind field $\overline{\mathbf{u}}_I$; the time-averaged wind field $\overline{\mathbf{u}}$; and the zonal components $\overline{u}_I$ (dashed line) and u (solid line) at $y = 0.44$ for a single time.

6. Discussion and conclusions

In this paper we have derived a balance model for planetary-scale motions in the presence of diabatic heating. Our work is heavily motivated by data assimilation: currently there are no balance relations that are valid in the tropics and capture Kelvin waves, and the relatively abundant mass observations are not used

effectively to constrain the errors in the wind field (Žagar 2012). This is a possible cause for the different estimates of tropical variability seen in different reanalyses, particularly for Kelvin waves.

We have applied the modified asymptotic approach of Warn et al. (1995) to derive a balance model from the shallow-water equations, using the ratio between the meridional and zonal length scales as a small parameter. The work here is an extension of CS13 to the diabatic regime, which is more relevant for the equatorial troposphere. To allow for the possibility of nonlinearity being important, we assumed the Froude number to be unity; in this case, the slow time scale can be interpreted as an advective time scale. We used a WTG-type scaling by assuming that the horizontal divergence is the same order as the mass source, but unlike the classical WTG model we did not assert this as the dominant balance in the mass equation.

Following CS13, we used the mass field variable h to describe the slow dynamics and tacitly assumed the wind field to be a function of the mass field via the balance relations. At leading order, the slow balance dynamics is characterized by a semigeostrophic balance between the zonal wind and meridional pressure gradient, while the meridional wind vanishes; the correction to the meridional wind can be found at the next order in the asymptotic expansion.

In principle the asymptotic method can be extended to obtain higher-order corrections, but in this paper we elected to focus our attention on the dynamics of the balance model. We integrated the balance model numerically with a mass source that varies periodically in time, and found the balance model to be in good agreement with the full SWE system for forcings that vary on long time scales (i.e., much longer than the inertial period); this indicates that the response of the full system is largely balanced. On the other hand, a reduction in the period of the forcing leads to a shortening in the wavelengths of the excited waves and the behavior of the balance model begins to diverge from that of the full system as the full system becomes more isotropic. More specifically, the periodic forcing excites spurious Rossby waves to the east of the forcing region, which is due to the model failing to properly model the rollover of the Rossby wave frequency at high wavenumbers (This problem could be mitigated by going to higher order in the expansion, see CS13.). Interestingly, even though the balance dynamics fail to reproduce the dynamics of the full system in the latter case, we found that the wind fields and mass field still satisfy the balance relations.

Additionally, we also compared the nonlinear longwave balance model derived here with the classical WTG model of Sobel et al. (2001). A major difference is that even though both models eliminate fast inertia–gravity waves, the WTG model does not admit free Kelvin waves: in the absence of heating, the WTG balance demands the divergence to vanish at leading order, and thereby filters out the Kelvin wave. Significant differences also exist from a diagnostic point of view. As most observations are of the temperature/mass field, the balance relationships are most often used operationally to infer the wind field from the mass field. With the nonlinear longwave balance model, the wind field can be determined completely by solving two linear equations. In contrast, for the WTG balance relations, inferring the wind field requires solving two partial differential equations, one of which is nonlinear. Furthermore, as diabatic heating is often not observed directly, this poses a major challenge for the WTG balance: without Q the horizontal divergence cannot be computed, which in turn makes inferring rotational wind impossible. In contrast, the zonal wind can be determined entirely using the mass field under the nonlinear longwave balance model, whereas the mass source Q is only required for the meridional wind (and even Q may be dispensed with if the time derivative of h is regarded as known). We should, however, emphasize that the nonlinear longwave balance model considered here does not necessarily contradict the assumptions of the WTG model, as it contains the seasonal planetary equatorial WTG model [derived by Majda and Klein (2003)] as a special limit.

Through several numerical experiments, we compared the use of WTG balance and the nonlinear longwave balance relations as a diagnostic tool for inferring horizontal divergence for a given mass field and mass source. We examined the Gill problem with both a stationary and time-varying mass source, and found that the nonlinear longwave balance relations performed better in all cases. The difference between the two balance relations is especially appreciable when the period of the forcing decreases.

We also considered an example where a steady forcing results in an unsteady solution; in this case the full SWE exhibits vortex shedding and excites MRG waves, neither of which are reproduced by the nonlinear longwave balance model; however, we have demonstrated that the time-averaged circulation is largely balanced, and the balance relations thus describe a ''guiding center'' about which the fast oscillations take place.

In the broader context, our work here indicates that the traditional longwave approximation (e.g., Gill 1980) can be interpreted as a balance model, since the smallness of the anisotropy parameter also indicates a separation of time scale. The nonlinear longwave balance model derived here unifies the equatorial longwave

dynamics with quasi nondivergence, which observations suggest is consistent with large-scale circulations in the tropics (Sardeshmukh and Hoskins 1987; Yano and Bonazzola 2009; Yano et al. 2009). Our model complements the WTG model, in the sense that the former is valid when the length scales are larger than the Rossby radius of deformation, whereas the latter is valid for scales smaller. We argue that the nonlinear longwave model is more relevant to planetary-scale dynamics, and that Kelvin waves should be regarded as slow in this regime. The distorted Rossby wave dynamics in the longwave model has previously been noted as a deficiency (Schubert and Masarik 2006; Schubert et al. 2009), and although this may hamper its use as a dynamical model, our results here indicate that the balance relations have a much wider regime of validity. This point is important as data assimilation procedures typically only apply balance constraints to large-scale structures. Furthermore, the vortex shedding experiment indicates that in the case where the motion is unbalanced, the balance relations can still be useful as they help determine the guiding center for the unbalanced motion.

Acknowledgments. We thank Jun-Ichi Yano and two other anonymous reviewers for their insightful comments on the paper. This research is funded by the Natural Sciences and Engineering Research Council. IC wishes to acknowledge additional financial support from the government of Ontario, Canada.

REFERENCES

Baldwin, M. P., and Coauthors, 2001: The quasi-biennial oscillation. *Rev. Geophys.,* **39,** 179–229, doi:10.1029/1999RG000073.

Boyd, J. P., 1980: The nonlinear equatorial Kelvin wave (in coastal currents of El Niño and Gulf of Guinea). *J. Phys. Oceanogr.,* **10,** 1–11.

Bretherton, C. S., and A. H. Sobel, 2003: The Gill model and the weak temperature gradient approximation. *J. Atmos. Sci.,* **60,** 451–460.

Chan, I. H., and T. G. Shepherd, 2013: Balance model for equatorial long waves. *J. Fluid Mech.,* **725,** 55–90.

Charney, J. G., 1963: A note on large-scale motions in the tropics. *J. Atmos. Sci.,* **20,** 607–608.

Delayen, K., and J. I. Yano, 2009: Is asymptotic non-divergence of the large-scale tropical atmosphere consistent with equatorial wave theories? *Tellus,* **61A,** 491–497.

Fraedrich, K., and J. L. McBride, 1989: The physical mechanism of CISK and the free-ride balance. *J. Atmos. Sci.,* **46,** 2642–2648.

Gill, A. E., 1980: Some simple solutions for heat-induced tropical circulation. *Quart. J. Roy. Meteor. Soc.,* **106,** 447–462.

——, and P. J. Philips, 1986: Nonlinear effects on heat-induced circulation of the tropical atmosphere. *Quart. J. Roy. Meteor. Soc.,* **112,** 69–91.

Heckley, W. A., and A. E. Gill, 1984: Some simple analytical solutions to the problem of forced equatorial long waves. *Quart. J. Roy. Meteor. Soc.,* **110,** 203–217.

Held, I. M., and A. Y. Hou, 1980: Nonlinear axially symmetric circulations in a nearly inviscid atmosphere. *J. Atmos. Sci.,* **37,** 515–533.

——, and B. J. Hoskins, 1985: Large-scale eddies and the general circulation of the troposphere. *Advances in Geophysics,* Vol. 28, Academic Press, 3–31.

Hoskins, B. J., and G.-Y. Yang, 2000: The equatorial response to higher-latitude forcing. *J. Atmos. Sci.,* **57,** 1197–1213.

Hsu, C. J., and R. A. Plumb, 2000: Nonaxisymmetric thermally driven circulations and upper-tropospheric monsoon dynamics. *J. Atmos. Sci.,* **57,** 1255–1276.

Kiladis, G. N., M. C. Wheeler, P. T. Haertel, K. H. Straub, and P. E. Roundy, 2009: Convectively coupled equatorial waves. *Rev. Geophys.,* **47,** RG2003, doi:10.1029/2008RG000266.

Majda, A. J., and R. Klein, 2003: Systematic multiscale models for the tropics. *J. Atmos. Sci.,* **60,** 393–408.

Matsuno, T., 1966: Quasi-geostrophic motions in the equatorial area. *J. Meteor. Soc. Japan,* **44,** 25–43.

Polvani, L. M., and A. H. Sobel, 2002: The Hadley circulation and the weak temperature gradient approximation. *J. Atmos. Sci.,* **59,** 1744–1752.

Ripa, P., 1982: Nonlinear wave wave interactions in a one-layer reduced-gravity model on the equatorial beta plane. *J. Phys. Oceanogr.,* **12,** 97–111.

Roundy, P. E., and W. M. Frank, 2004: A climatology of waves in the equatorial region. *J. Atmos. Sci.,* **61,** 2105–2132.

Sardeshmukh, P. D., and B. J. Hoskins, 1985: Vorticity balances in the tropics during the 1982-83 El Niño–Southern Oscillation event. *Quart. J. Roy. Meteor. Soc.,* **111,** 261–278.

——, and ——, 1987: On the derivation of the divergent flow from the rotational flow: The χ-problem. *Quart. J. Roy. Meteor. Soc.,* **113,** 339–360.

Schubert, W. H., and M. T. Masarik, 2006: Potential vorticity aspects of the MJO. *Dyn. Atmos. Oceans,* **42,** 127–151.

——, L. G. Silvers, M. T. Masarik, and A. O. Gonzalez, 2009: A filtered model of tropical wave motions. *J. Adv. Model. Earth Syst.,* **1,** 3, doi:10.3894/JAMES.2009.1.3.

Sobel, A. H., J. Nilsson, and L. M. Polvani, 2001: The weak temperature gradient approximation and balanced tropical moisture waves. *J. Atmos. Sci.,* **58,** 3650–3665.

Stevens, D. E., H.-C. Kuo, W. H. Schubert, and P. E. Ciesielski, 1990: Quasi-balanced dynamics in the tropics. *J. Atmos. Sci.,* **47,** 2262–2273.

Van Kampen, N. G., 1985: Elimination of fast variables. *Phys. Rep.,* **124,** 69–160.

Van Tuyl, A. H., 1986: Advective influences on forced tropical motions. *J. Atmos. Sci.,* **43,** 141–161.

Wallace, J. M., and V. E. Kousky, 1968: Observational evidence of Kelvin waves in the tropical stratosphere. *J. Atmos. Sci.,* **25,** 900–907.

Warn, T., O. Bokhove, T. G. Shepherd, and G. K. Vallis, 1995: Rossby number expansions, slaving principles, and balance dynamics. *Quart. J. Roy. Meteor. Soc.,* **121,** 723–739.

Webster, P. J., 1972: Response of the tropical atmosphere to local, steady forcing. *Mon. Wea. Rev.,* **100,** 518–541.

Wheeler, M., and G. N. Kiladis, 1999: Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber-frequency domain. *J. Atmos. Sci.,* **56,** 374–399.

Yano, J.-I., and M. Bonazzola, 2009: Scale analysis for large-scale tropical atmospheric dynamics. *J. Atmos. Sci.,* **66,** 159–172.

——, S. Mulet, and M. Bonazzola, 2009: Tropical large-scale circulations: Asymptotically nondivergent? *Tellus,* **61A,** 417–427.

Žagar, N., 2012: Multivariate data assimilation in the tropics by using equatorial waves. *Pure Appl. Geophys.,* **169,** 367–379.

Zhang, C., 1993: Laterally forced equatorial perturbations in a linear model. Part II: Mobile forcing. *J. Atmos. Sci.,* **50,** 807–821.

——, and P. J. Webster, 1992: Laterally forced equatorial perturbations in a linear model. Part I: Stationary transient forcing. *J. Atmos. Sci.,* **49,** 585–607.

Zhou, B., and A. H. Sobel, 2006: Nonlinear shallow-water solutions using the weak temperature gradient approximation. *Theor. Comput. Fluid Dyn.,* **20,** 469–484.