Near-Geostrophic Approximations of the Spherical Shallow-Water Equations

Jeffrey J. Early * R. M. Samelson and Dudley B. Chelton

College of Oceanic and Atmospheric Sciences, Corvallis, Oregon

* *Corresponding author address:* Jeffrey J. Early, College of Oceanic and Atmospheric Sciences, 104 COAS Admin Bldg, Oregon State University, Corvallis, OR 97331-5503, USA.

E-mail: jeffrey@jeffreyearly.com

ABSTRACT

We use a novel technique for evaluating scales of motion in order find an appropriate model for altimeter tracked mesoscale eddies. Starting from the spherical shallow water equations and assuming geostrophic dominance, we derive a potential vorticity conservation law in terms of all four non-dimensional parameters inherent in the equations while retaining the spherical geometry. The resulting equation reduces to other existing geostrophic theories, such as quasi-geostrophy and the Flierl-Petviashvili equation, by assuming a precise relationship between the non-dimensional parameters. However, by retaining freedom in the parameters we can determine at what scales the various theories remain valid. We find that a new extension to the FP equation is required to describe the mid-latitude mesoscale eddies.

1. Introduction

Fluid motions in the ocean span a range of scales from millimeters to thousands of kilometers and operate on time scales from seconds to thousands of years. Even by neglecting all thermodynamic effects and other complicating interactions like the atmosphere or biological input, we still stand little chance of finding solutions to our mathematical description without further simplification. A typical approach to simplification is to take a sufficiently general set of equations like the Navier-Stokes equations and then use our knowledge of some observed phenomenon to apply additional constraints. For example, by taking J. Scott Russell's observations of solitons, we can restrict the Navier Stokes equations to one-dimensional, irrotational, small-amplitude surface waves to find the Korteweg de Vries equation Ablowitz and Clarkson (1991). If our assumptions about the scales and physics of the observations are sufficiently accurate, then the simplified equations should admit solutions similar to the observations. Of course, a number of things can go wrong during such a derivation, including starting with equations already overly simplified, using bad assumptions about the observations, finding an equation still too difficult to solve or even finding an equation so simple that it no longer describes the observations. In all of these cases, we must go back to earlier states of the derivation and reassess the assumptions made, in hopes of deriving a better theory.

This general process is exactly the pattern followed for satellite altimetry observations of oceanic mesoscales features, which were first compared to linear Rossby waves and shown to differ from theoretical predictions Chelton and Schlax (1996). This discovery was followed by attempts to modify the classical linear theory (e.g., Killworth et al. (1997)) only to have subsequent observations change the interpretation of the observations from linear waves to non-linear eddies Chelton et al. (2007). Amongst these observations it was found that these eddies are characterized by a strong degree of non-linearity as measured by three nondimensional parameters involving eddy height, length scale and typical fluid speed. These nonlinearity parameters are exactly the information we require to modify our assumptions and find a new theory to explain the observations.

In the original interpretation of the satellite altimetry observations, a linearized form of quasi-geostrophic (QG) theory that allows the superposition of linear Rossby waves (Pedlosky (1987)) is employed to explain the observations. However, as it now believed that the observations show features that are more eddy-like than wave-like (Chelton, et. al., in preparation), the fully non-linear quasi-geostrophic equations are a much more likely candidate theory. On the other hand, given the strong fluid velocities, perhaps a second order extension to QG theory is required? Alternatively, the strong height non-linearity suggests QG needs to be extended to include larger amplitudes, as in the Flierl-Petviashvili equation? Or, perhaps effects from neglecting the earth's curvature are required, owing to the long eddy length scales? Any one or all of these cases may be a possibility and an appropriate theory should demonstrate its validity over the *range* of observed scales.

In the usual asymptotic derivation of QG, the non-dimensional parameters admitted by the primitive equations are fixed relative to single parameter, which is typically the Rossby number defined as $\epsilon = U/fL$ where U and L are the characteristic fluid speed and length scale, respectively, and f is the Coriolis parameter. The dependent variables are then expanded as a perturbation series with respect to ϵ and in the limit as $\epsilon \mapsto 0$ only the leading order remains. Quasi-geostrophic theory stems from one particular choice of relationships amongst the non-dimensional parameters. However, other choices are possible. In Williams and Yamagata (1984) a 5x5 matrix illustrating different choices between non-dimensional parameters shows the corresponding equations ultimately resulting from the perturbation expansions. In a similar vein, Charney and Flierl (1981) sketch a continuous two dimensional plot of nonlinear parameters showing different regions of behavior requiring different theoretical descriptions.

In this paper we expand on these ideas by showing how a potential vorticity equation can be derived from the spherical shallow water equations without fixing the non-dimensional parameters *a priori*. The resulting equation can then be reduced to other theories, such as quasi-geostrophy and the Flierl-Petviashivilli equation, by assuming a precise relationship between the parameters. However, by retaining the freedom in the parameters we can quickly deduce which terms in the equation are necessary for an accurate description of the observed scales. Conversely, we are also able to determine at what scales the various theories remain valid and at what scales their assumptions are violated. To specifically address the non-linear effects of height, fluid speed, and length scale, the spherical shallow water equations are a reasonable starting point, although this does neglect other potentially important effects. We show that the quasi-geostrophic potential vorticity equation is not a valid equation for the observed scales of the eddies and that the Flierl-Petviashvili equation and an extended FP equation are more suitable models.

This approach removes some of the burden of restating the assumptions when finding a new theory to describe observations. However, these ideas are not necessarily widely applicable and owe their success to the particular form of the primitive equations and the wide range of scales over which geostrophy dominates other effects for mesoscale and large-scale oceanography. Furthermore, we consider only the single layer shallow water equations in spherical coordinates, for which a number of (conceivably invalid) assumptions have already been made to reduce the complexity of the equations.

2. Balance Dynamics

Consider one possible formulation of the primitive equations,

$$u_t + (uu_x + vu_y) - fv = -\,gh_x \tag{1}$$

$$v_t + (uv_x + vv_y) + fu = -\,gh_y \tag{2}$$

$$h_t + (uh_x + vh_y) = -\,h\,(u_x + v_y)\,. \tag{3}$$

Equations (1) and (2) represent the horizontal momentum equations and (3) represents the continuity equation where (u, v) are the x and y fluid speeds and h is the fluid height. The fluid height h is also thought of as an equivalent depth scale fluid height for an equivalent barotropic model. Without further simplification, these coupled partial differential equations are difficult to solve. However, a key principle in oceanography is that mesoscale and large-scale features have pressure fields and velocity fields largely in geostrophic balance. These features evolve on a 'slow' time scale, while the effects of inertia-gravity waves operate on a 'fast' time scale and can often be ignored Lighthill (1952); Ford et al. (2002). The value in this principle is that it allows for a simplification of primitive equations (1)-(3). Rather than solving three partial differential equations for all three variables simultaneously, the equation can be reduced to a set of *balance relations* and *balance dynamics* Warn et al. (1995) Vallis (1996). For example, separating the fluid height into a static depth and surface perturbation $h = D + \eta$, the balance relations could be taken as

$$\begin{aligned} v =& \frac{g}{f}\eta_x \\ u =& -\frac{g}{f}\eta_y \end{aligned} \tag{4}$$

while the balance dynamics might be the quasi-geostrophic potential vorticity equation (QG-PVE),

$$\nabla^2 \eta_t - \eta_t + J(\eta, \nabla^2 \eta) = 0 \tag{5}$$

where $J(a, b) = a_x b_y - a_y b_x$. The evolution of the system is therefore entirely determined by the evolution of the height field in equation (5) and the fluid velocity (u, v) follows diagnostically from (4).

The usual approach to determining the balance relations and balance dynamics is to expand the variables in a perturbation series as a function of some small parameter. For example, take the nondimensionalized, linear f-plane primitive equations as

$$v = \eta_x + \epsilon u_t \tag{6}$$

$$u = -\eta_y - \epsilon v_t \tag{7}$$

$$\epsilon\eta_t + \epsilon(u\eta_x + v\eta_y) = -(1+\epsilon\eta)(u_x + v_y) . \tag{8}$$

where $\epsilon \ll 1$. Given that $\epsilon \ll 1$, we can write the three field variables as a perturbation series Pedlosky (1987),

$$\begin{aligned} v &= v^0 + \epsilon v^1 + \epsilon^2 v^2 + O\left(\epsilon^3\right) \\ u &= u^0 + \epsilon u^1 + \epsilon^2 u^2 + O\left(\epsilon^3\right) \\ \eta &= \eta^0 + \epsilon \eta^1 + \epsilon^2 \eta^2 + O\left(\epsilon^3\right) . \end{aligned} \tag{9}$$

If equation (9) is inserted into the primitive equations (6)-(8) and terms of the same order in ϵ are collected, expressions for u^i, v^i and η^i are found. To find a potential vorticity equation, take the curl of the momentum equations at each order of ϵ individually. The resulting expression depends on η^0 and the ageostrophic component of the momentum equations (u^1, v^1). The ageostrophic component can be eliminated in favor of the lower order η^0 to find a potential vorticity equation dependent only on η_0. This is the balance dynamics. The resulting fluid velocities are determined from the leading order values in the expansion, $u_0 = -\eta^0_y$ and $v_0 = \eta^0_x$ and provide the balance relations.

Aside from the limitations of this approach due to fixing the non-dimensional parameters *a priori* that were highlighted in the introduction, Warn et al. (1995) also showed that expanding η often leads to secularities in higher order terms. To avoid these secularities, it is suggested to leave the variable η unexpanded. Interestingly, this turns out to be a similar approach as the iterated models in Allen (1993). The idea is to substitute equation (6) into (7) and vice versa so that,

$$v = \eta_x - \epsilon \eta_{yt} + \epsilon^2 v_{tt} \tag{10}$$

$$u = -\eta_y - \epsilon \eta_{xt} - \epsilon^2 u_{tt}. \tag{11}$$

Equations (10) and (11) are still exact, but now depend *only* on η through order ϵ. Repeated application of this iterative process allow the fluid speeds to be determined by η though arbitrary orders of ϵ. By truncating at order n of ϵ one obtains an order $n-1$ balance relation for u and v in terms of the height η. Of course, without balance dynamics of similar order, the balance relations aren't of much use.

The approach taken here is to find an equation for the balance dynamics from the potential vorticity conservation law of the spherical shallow water equations. Using the same iterative approach as for the momentum equation, the potential vorticity law is similarly used to determine the balance dynamics to arbitrary order. A key feature to this approach is that a single, general equation can be derived without fixing the non-dimensional parameters relative to one another, unlike perturbation expansions.

3. Field Equations

The shallow water equations in spherical coordinates Gill (1982) include the momentum equations,

$$\frac{\partial u}{\partial t} + \frac{u}{R\cos\theta}\frac{\partial u}{\partial \phi} + \frac{v}{R}\frac{\partial u}{\partial \theta} = \left(\frac{u}{R}\tan\theta + f\right) v - \frac{g}{R\cos\theta}\frac{\partial h}{\partial \phi} \tag{12}$$

$$\frac{\partial v}{\partial t} + \frac{u}{R\cos\theta}\frac{\partial v}{\partial \phi} + \frac{v}{R}\frac{\partial v}{\partial \theta} = -\left(\frac{u}{R}\tan\theta + f\right) u - \frac{g}{R}\frac{\partial h}{\partial \theta} \tag{13}$$

and with the addition of continuity,

$$\frac{\partial h}{\partial t} + \frac{u}{R\cos\theta}\frac{\partial h}{\partial \phi} + \frac{v}{R}\frac{\partial h}{\partial \theta} = -\frac{h}{R\cos\theta}\left[\frac{\partial u}{\partial \phi} + \frac{\partial}{\partial \theta}(v\cos\theta)\right]. \tag{14}$$

Equations (12)-(14) have a number of conserved quantities including angular momentum, energy and potential vorticity Vallis (2006). The conservation of energy can be written as

$$\frac{\partial E}{\partial t} + \nabla\cdot\mathbf{S} = 0 \tag{15}$$

where $E = \frac{1}{2}\left(gh + u^2 + v^2\right)$ and $\mathbf{S} = \frac{1}{2}\left(2gh + u^2 + v^2\right)$.

If we cross-differentiate (12) and (13) and use (14), then we can show that potential vorticity is conserved. Namely that,

$$\frac{d}{dt}\left[\frac{1}{h}\left(f + \frac{1}{R\cos\theta}\left(\frac{\partial v}{\partial \phi} - \frac{\partial}{\partial \theta}(u\cos\theta)\right)\right)\right] = 0, \tag{16}$$

where $\frac{d}{dt} = \frac{\partial}{\partial t} + \frac{u}{R\cos\theta}\frac{\partial}{\partial \phi} + \frac{v}{R}\frac{\partial}{\partial \theta}$, or equivalently,

$$\frac{d}{dt}\left[\frac{1}{h}\left(f + \frac{u}{R}\tan\theta + \frac{1}{R\cos\theta}\frac{\partial v}{\partial \phi} - \frac{1}{R}\frac{\partial u}{\partial \theta}\right)\right] = 0. \tag{17}$$

4. Coordinates

Before proceeding to derive an equation for the balance dynamics, it is worth considering the coordinates that we're working in and the associated gradient and laplacian operators. One of the primary objectives of this study is to assess the validity of approximating the equations of motion to a cartesian plane, so we need to carefully consider our definition of coordinates.

If we define the (x', y', z') coordinate system with origin at the center of the earth in an inertial frame, then in terms of the coordinates (ϕ, θ, r) (denoted here as geographer's coordinates)

$$\begin{aligned} x' =& (R+r)\cos\theta\cos(\phi+\omega t), \\ y' =& (R+r)\cos\theta\sin(\phi+\omega t), \\ z' =& (R+r)\sin\theta, \\ t =& t. \end{aligned} \tag{18}$$

Restricting ourselves to motion on the geoid, we can ignore changes in r and neglect the small metric terms Gill (1982). The gradient operator in the geographer's coordinates is

$$\nabla\zeta = \left(\frac{1}{R}\frac{\partial\zeta}{\partial\theta}, \frac{1}{R\cos\theta}\frac{\partial\zeta}{\partial\phi}\right) \tag{19}$$

and the laplacian is

$$\nabla^2\zeta = \frac{1}{R^2\cos\theta}\frac{\partial}{\partial\theta}\left(\frac{\partial\zeta}{\partial\theta}\cos\theta\right) + \frac{1}{R^2\cos^2\theta}\frac{\partial^2\zeta}{\partial\phi^2}. \tag{20}$$

We have to be careful when trying to write the above operators in typical oceanographer's coordinates (a local Cartesian plane with (x, y, z)) because in both cases there is a curvature term that is often ignored. In order to create coordinates resembling a local Cartesian plane, oceanographers typically define the following horizontal coordinates,

$$\begin{aligned} x =& R\phi \cos\theta_0, \\ y =& R(\theta - \theta_0), \end{aligned} \tag{21}$$

where θ_0 is the latitude of the origin of the Cartesian plane. We need to use these coordinates for derivatives, so by the chain rule we find that,

$$\frac{\partial}{\partial \phi} = R\cos\theta_0 \frac{\partial}{\partial x} \tag{22}$$

$$\frac{\partial}{\partial \theta} = R\frac{\partial}{\partial y}. \tag{23}$$

Writing the Laplacian in these coordinates gives us,

$$\nabla^2\zeta = \frac{\cos^2\theta_0}{\cos^2\theta}\frac{\partial^2\zeta}{\partial x^2} + \frac{\partial^2\zeta}{\partial y^2} - \frac{\tan\theta}{R}\frac{\partial\zeta}{\partial y}. \tag{24}$$

The coefficient of the x derivative terms stays close to 1 until motions deviate too far north or south of the cartesian origin. Similarly, the additional $\tan\theta$ dependent curvature term can remain fairly small if the distances being considered don't get too large.

In this study, we will write the Laplacian by keeping the additional curvature term separate. For the remainder of this paper the following notation will be used,

$$\nabla^2\zeta = \frac{\cos^2\theta_0}{\cos^2\theta}\frac{\partial^2\zeta}{\partial x^2} + \frac{\partial^2\zeta}{\partial y^2} - \frac{\tan\theta}{R}\frac{\partial\zeta}{\partial y} \tag{25}$$

$$= \nabla_H^2\zeta - \frac{\tan\theta}{R}\frac{\partial\zeta}{\partial y} \tag{26}$$

and therefore we've implicitly defined

$$\nabla_H^2\zeta = \frac{\cos^2\theta_0}{\cos^2\theta}\frac{\partial^2\zeta}{\partial x^2} + \frac{\partial^2\zeta}{\partial y^2}. \tag{27}$$

5. Nondimensionalization

The objective of this section is to nondimensionalize the oceanographer's coordinates so that we can write the shallow water equations in a form suitable for the iterative scheme.

a. *Coordinates*

Simply by inspection we can observe all of the dimensional parameters in the shallow water equations (12)-(14). We have t, x, and y for the independent variables, u, v, and h for the dependent variables and additionally R, g and 2ω appear as constants. Nondimensionalizing requires that we separate the dimensional part of the variables. We do this in the following fashion,

$$t = T\bar{t}, \quad x = L\bar{x}, \quad y = L\bar{y}, \quad u = U\bar{u}, \quad v = U\bar{v}, \quad h = D + N_0\eta. \tag{28}$$

The non-dimensional oceanographer's coordinates are now,

$$\bar{x} = \frac{R}{L}\phi\cos\theta_0, \tag{29}$$

$$\bar{y} = \frac{R}{L}(\theta - \theta_0), \tag{30}$$

and so by the chain rule we find that,

$$\frac{\partial}{\partial\phi} = \frac{R}{L}\cos\theta_0\frac{\partial}{\partial\bar{x}} \tag{31}$$

$$\frac{\partial}{\partial\theta} = \frac{R}{L}\frac{\partial}{\partial\bar{y}}. \tag{32}$$

Important to note here is that we made the assumption that x and y scale similarly (they both use L as their dimensional parameter) and that u and v do the same (using U for dimensionality). The result of doing this means that we essentially reduced the number of dimensional parameters in the problem from 9 down to 7. However, we added an additional dimension by separating h into a static depth D and a relatively small variation N_0. We are therefore left with 8 dimensional parameters in the problem. To be explicit, the dimensional parameters are now T, L, U, D, N_0, R, g and 2ω.

The Buckingham-Pi theorem Bluman and Anco (2002) states that because we have only two fundamental physical units, length and time, and eight dimensional parameters there are six dimensionless parameters we can construct. However, as soon as we approximate the equations on a cartesian plane, the latitude, θ_0, also counts as a non-dimensional parameter. There are an infinite number of ways of constructing these parameters, but we will use the 'natural' scaling that arises in our particular problem.

We will use one unusual convention for writing a non-dimensional form of the trigonometric terms. We write $\overline{\sin\theta} = \frac{\sin\theta}{\sin\theta_0}$ to indicate a normalized form of trig functions. Because $\theta = \theta_0 + \frac{L}{R}\bar{y}$ this means that $\overline{\sin\theta} \approx 1$ if $\frac{L}{R} \ll 1$, which can often be the case. This becomes a convenient way to keep track of which terms are affected by the spherical geometry of the problem and at what order in our expansion the spherical effects start to become significant.

b. Field Equations

To write equations (12)-(14) in a non-dimensionalized form, use the horizontal derivatives (31)-(31), substitute in the non-dimensional scalings (28) and then divide the equations by $U \cdot 2 \cdot \omega$. Traditionally one would also divide by a factor of $\sin\theta_0$, but that ultimately makes the analysis much more difficult. With mild rearranging, the shallow water equations become,

$$\bar{v} = \frac{G}{\sin\theta\overline{\cos\theta}}\frac{\partial\eta}{\partial\bar{x}} + \epsilon_T\cdot\frac{1}{\sin\theta}\frac{\partial\bar{u}}{\partial\bar{t}} + \epsilon_R\cdot\frac{1}{\sin\theta}\left[\frac{\bar{u}}{\overline{\cos\theta}}\frac{\partial\bar{u}}{\partial\bar{x}} + \bar{v}\frac{\partial\bar{u}}{\partial\bar{y}}\right] - \epsilon_\beta\cdot\frac{\bar{u}\,\bar{v}}{\cos\theta}, \tag{33}$$

$$\bar{u} = -\frac{G}{\sin\theta}\frac{\partial\eta}{\partial\bar{y}} - \epsilon_T\cdot\frac{1}{\sin\theta}\frac{\partial\bar{v}}{\partial\bar{t}} - \epsilon_R\cdot\frac{1}{\sin\theta}\left[\frac{\bar{u}}{\overline{\cos\theta}}\frac{\partial\bar{v}}{\partial\bar{x}} + \bar{v}\frac{\partial\bar{v}}{\partial\bar{y}}\right] - \epsilon_\beta\cdot\frac{\bar{u}^2}{\cos\theta}. \tag{34}$$

and

$$\frac{\epsilon_H\epsilon_T}{\epsilon_R}\eta_{\bar{t}} + \epsilon_H\frac{\bar{u}}{\overline{\cos\theta}}\frac{\partial\eta}{\partial\bar{x}} + \epsilon_H\bar{v}\frac{\partial\eta}{\partial\bar{y}} = -\frac{1+\epsilon_H\eta}{\overline{\cos\theta}}\left(\frac{\partial\bar{u}}{\partial\bar{x}} + \frac{\partial}{\partial\bar{y}}\left(\bar{v}\overline{\cos\theta}\right)\right), \tag{35}$$

where

$$\boxed{\epsilon_T = \frac{1}{2\omega T},\quad \epsilon_R = \frac{U}{2\omega L},\quad \epsilon_\beta = \frac{U}{2\omega R}\quad G = \frac{gN_0}{2\omega UL},\quad \epsilon_H = \frac{N_0}{D}} \tag{36}$$

The first two parameters, ϵ_T and ϵ_R are the commonly used Rossby numbers without the extra factor of $\sin\theta_0$. The third parameter, ϵ_β is related to the usual parameter β_0 by $\beta_0 = \epsilon_\beta\frac{(2\omega)^2\cos\theta_0}{U}$. The G parameter is a geostrophic pressure parameter. The continuity equation (35) includes ϵ_H which measures the surface perturbation relative to the equivalent depth. The sixth parameter (required by the Buckingham-Pi theorem) can be written as the aspect ratio of the ocean, namely the depth scale over the length scale $\frac{D}{L}$. It does not naturally appear in the equations that we started with, but was implicitly already used to get the hydrostatic balance. Any of the other standard non-dimensional parameters used for these equations can easily be recovered in terms of (36). For example, the Burger parameter $F = \frac{L^2}{L_R^2}$ can be constructed from (36) with $F = \frac{\epsilon_H}{G\epsilon_R}\sin^2\theta_0$.

Two of the coefficients in equations (33)-(35) can be varied to match a third by choosing the units of time and length. In a typical derivation of quasi-geostrophy the geostrophic parameter G is taken to be 1 and ϵ_T is set to equal ϵ_R Pedlosky (1987). However, it's important to note that G could just as well be set equal to ϵ_β, and ϵ_R set to 10^5 or any other value. This is the central idea to nondimensionalization and is done without loss of generality, provided we only use the two degrees of freedom in dimensions. Of course, the values of these parameters might as well be chosen to reflect our assumptions about the resulting solution. So in the case of near geostrophic flows, while setting $G = 1$ may represent the idea of purely geostrophic motion such that $u = -\frac{g}{f_0}\eta_x$, the solution is is not restricted to this assumption owing to the freedom in dimensionality. Using what we think will be typical values for ϵ_T and ϵ_R, we may decide that $\epsilon_T = \epsilon_R$, or instead that $\epsilon_T^{1.5} = \epsilon_R$. Doing this is still without loss of generality, but now provides a concrete way of ordering the relative magnitude of the terms based on our *assumptions* of the resulting solutions. This is how one typically proceeds with a perturbation analysis, but by continuing the derivation without fixing the units, it becomes easier at a later stage to estimate the relative sizes of the terms. Specifically, we will also use the freedom in G to later estimate the range at which departures from pure geostrophy ($G = 1$) are still valid.

It's also important to note that the ratio of length scales $\frac{L}{R}$ can be written as $\frac{\epsilon_\beta}{\epsilon_R}$. This means that we can write $\frac{\partial}{\partial \bar{y}} = \frac{\epsilon_\beta}{\epsilon_R}\frac{\partial}{\partial \theta}$.

c. *Potential Vorticity*

For notational convenience take $\omega_a = \left(f + \frac{u}{R}\tan\theta + \frac{1}{R\cos\theta}\frac{\partial v}{\partial \phi} - \frac{1}{R}\frac{\partial u}{\partial \theta}\right)$ so that we can rewrite the potential vorticity conservation law (17),

$$
\begin{aligned}
\frac{d}{dt}\left(\frac{\omega_a}{h}\right) &= 0 \\
\frac{1}{h}\frac{d\omega_a}{dt} - \frac{\omega_a}{h^2}\frac{dh}{dt} &= 0 \\
\frac{d\omega_a}{dt} - \omega_a\frac{d}{dt}\ln h &= 0.
\end{aligned} \tag{37}
$$

It is this last form of the equation that we will ultimately find most convenient. Nondimensionalizing we find

$$\omega_a = 2\omega \cdot \left[\sin\theta + \epsilon_R \left(\frac{1}{\cos\theta}\frac{\partial \bar{v}}{\partial \bar{x}} - \frac{\partial \bar{u}}{\partial \bar{y}}\right) + \epsilon_\beta \tan\theta \bar{u}\right] \tag{38}$$

$$h = D\left[1 + \epsilon_H \eta\right] \tag{39}$$

$$\frac{d}{dt} = 2\omega \cdot \left[\epsilon_T \frac{\partial}{\partial \bar{t}} + \epsilon_R \left(\frac{\bar{u}}{\cos\theta}\frac{\partial}{\partial \bar{x}} + \bar{v}\frac{\partial}{\partial \bar{y}}\right)\right]. \tag{40}$$

Now taking equations (38)-(40) and substituting them into equation (37) we see that the non-dimensionalized form of potential vorticity conservation is

$$0 = \left[\epsilon_T \frac{\partial}{\partial \bar{t}} + \epsilon_R \left(\frac{\bar{u}}{\cos\theta}\frac{\partial}{\partial \bar{x}} + \bar{v}\frac{\partial}{\partial \bar{y}}\right)\right]\left[\sin\theta + \epsilon_R \left(\frac{1}{\cos\theta}\frac{\partial \bar{v}}{\partial \bar{x}} - \frac{\partial \bar{u}}{\partial \bar{y}}\right) + \epsilon_\beta \tan\theta \bar{u}\right]$$
$$- \left[\sin\theta + \epsilon_R \left(\frac{1}{\cos\theta}\frac{\partial \bar{v}}{\partial \bar{x}} - \frac{\partial \bar{u}}{\partial \bar{y}}\right) + \epsilon_\beta \tan\theta \bar{u}\right]\left[\epsilon_T \frac{\partial}{\partial \bar{t}} + \epsilon_R \left(\frac{\bar{u}}{\cos\theta}\frac{\partial}{\partial \bar{x}} + \bar{v}\frac{\partial}{\partial \bar{y}}\right)\right] \ln\left[1 + \epsilon_H \eta\right]. \tag{41}$$

6. Near-Geostrophic Potential Vorticity Equation

The guiding assumption is that the flow is nearly in geostrophic balance, so that $\epsilon_T, \epsilon_R, \epsilon_\beta \ll G$. We will use the notation $O(\epsilon^n)$ to indicate $O(\epsilon_T^i \epsilon_R^j \epsilon_\beta^k)$ where $n = i + j + k$. Although ϵ_H is generally small for the scales that we are interested in, we do not need to assume that $\epsilon_H \ll G$ in order to proceed and therefore will not include it as part of the $O(\epsilon^n)$ notation.

By substituting equations (33) and (34) into (41) twice, we will derive a single scalar equation in η through $O(\epsilon^2)$, with combinations of η, $\bar{u}$ and $\bar{v}$ at order $O(\epsilon^3)$. This equation will still be exact and contain no errors, just as in equations (10) and (11).

The tedious computation was completed by hand and checked using a symbolic manipulator.

a. The Complete Equation

$$
\begin{aligned}
0 =&\epsilon_\beta G\cdot && \frac{\cos\theta}{\sin\theta}\frac{\eta_{\bar{x}}}{\cos\theta} \\
&+\epsilon_T\epsilon_R G\cdot && \frac{1}{\sin\theta}\nabla^2\eta_{\bar{t}} \\
&-\epsilon_T\epsilon_\beta G\cdot && \frac{1+\cos^2\theta}{\sin^2\theta\cos\theta}\eta_{\bar{t}\bar{y}} \\
&+\epsilon_R^2 G^2\cdot && \frac{1}{\sin^2\theta}\frac{1}{\cos\theta}\left[-\frac{\eta_{\bar{y}}\eta_{\bar{x}\bar{x}\bar{x}}}{\cos^2\theta}+\eta_{\bar{x}}\eta_{\bar{y}\bar{y}\bar{y}}-\eta_{\bar{y}}\eta_{\bar{x}\bar{y}\bar{y}}+\frac{\eta_{\bar{x}}\eta_{\bar{x}\bar{x}\bar{y}}}{\cos^2\theta}\right] \\
&+\epsilon_R\epsilon_\beta G^2\cdot && \frac{\cos\theta}{\sin^3\theta}\left[-3\frac{\eta_{\bar{x}}\eta_{\bar{y}\bar{y}}}{\cos\theta}+\frac{1+\cos^2\theta}{\cos^2\theta}\frac{\eta_{\bar{y}}\eta_{\bar{x}\bar{y}}}{\cos\theta}+\frac{2-3\cos^2\theta}{\cos^2\theta}\frac{\eta_{\bar{x}}\eta_{\bar{x}\bar{x}}}{\cos^3\theta}\right] \\
&-\epsilon_R\epsilon_\beta G\cdot && \frac{\eta_{\bar{y}\bar{y}}}{\cos\theta} \\
&+\epsilon_\beta^2 G^2\cdot && \frac{\cos^4\theta+3\cos^2\theta-1}{\sin^4\theta\cos^2\theta}\frac{\eta_{\bar{x}}\eta_{\bar{y}}}{\cos\theta} \\
&+\epsilon_\beta^2 G\cdot && \frac{1}{\sin\theta}\eta_{\bar{y}} \\
&-\epsilon_H\epsilon_T\cdot && \sin\theta\,\eta_{\bar{t}}(1+\epsilon_H\eta)^{-1} \\
&+\epsilon_H\epsilon_T\epsilon_R G\cdot && \frac{1}{\sin\theta}\left[\frac{1}{2}\frac{\partial}{\partial\bar{t}}\left(\frac{\eta_{\bar{x}}^2}{\cos^2\theta}+\eta_{\bar{y}}^2\right)-\eta_{\bar{t}}\cdot\nabla^2\eta\right](1+\epsilon_H\eta)^{-1} \\
&+\epsilon_H\epsilon_T\epsilon_\beta G\cdot && \frac{\eta_{\bar{y}}\eta_{\bar{t}}}{\sin^2\theta\cos\theta}(1+\epsilon_H\eta)^{-1} \\
&-\epsilon_H\epsilon_R^2 G^2\cdot && \frac{1}{\sin^2\theta}\frac{1}{\cos\theta}\left[\eta_{\bar{x}}\eta_{\bar{y}}\left(\frac{\eta_{\bar{x}\bar{x}}}{\cos^2\theta}-\eta_{\bar{y}\bar{y}}\right)+\eta_{\bar{x}\bar{y}}\left(\eta_{\bar{y}}^2-\frac{\eta_{\bar{x}}^2}{\cos^2\theta}\right)\right](1+\epsilon_H\eta)^{-1} \\
&-\epsilon_H\epsilon_R\epsilon_\beta G^2\cdot && \frac{\cos\theta}{\sin^3\theta}\frac{\eta_{\bar{x}}}{\cos\theta}\left[\frac{\cos 2\theta}{\cos^2\theta}\frac{\eta_{\bar{x}}^2}{\cos^2\theta}+\eta_{\bar{y}}^2\right](1+\epsilon_H\eta)^{-1}+O(\epsilon^3)
\end{aligned}
\tag{42}
$$

Equation (42) is still *exact*. We made no approximations from the original potential vorticity equation (17) and have simply hidden the remaining finite (but too numerous to print!) number of terms in the $O(\epsilon^3)$ rather than write them out explicitly. Unless we perform an additional iteration of inserting equations (33) and (34) into (42), there are no $O(\epsilon^4)$ terms. In order to find an equation for the balance dynamics solely in terms of η, we will need to drop at a minimum the $O(\epsilon^3)$ terms because they contain explicit reference to $\bar{u}$ and $\bar{v}$. For this reason it, it will be necessary to explicitly consider a few of the $O(\epsilon^3)$ terms.

b. Bounding the Cubic Terms

To determine the magnitude of the $O(\epsilon^3)$ terms, first consider the relative sizes of ϵ_T, ϵ_R, and ϵ_β. The ratio of parameters ϵ_β and ϵ_R is the factor $\frac{L}{R}$. Even an eddy with length scale 600 km will be an order of magnitude smaller than the radius of the earth and therefore we can consistently regard $\epsilon_\beta \ll \epsilon_R$. On the other hand, the condition that $\epsilon_T < \epsilon_R$ requires that $L/T < U$, a condition not supported by observation Chelton et al. (2007) if we take L/T to be the baroclinic wave speed. Closer to the equator ϵ_T tends to dominate while at mid-latitudes ϵ_R tends to dominate. This means that the largest $O(\epsilon^3)$ term that we can construct would be either an ϵ_R^3 or ϵ_T^3 term. However, equation (42) contains no ϵ_T^3 nor $\epsilon_T^2 \epsilon_R$ term and so depending on the latitude, either the ϵ_R^3 or $\epsilon_T \epsilon_R^2$ term will be the largest $O(\epsilon^3)$ term.

The $O(\epsilon_R^3)$ terms look like $\frac{\epsilon_R^3 G}{\sin^2\theta}\left[\frac{2\eta_{\bar{x}}\bar{u}_{\bar{x}}\bar{v}_{\bar{x}\bar{t}}}{\cos^3\theta} + ...\right]$ with the same parameters of $\frac{G\epsilon_R^3}{\sin^2\theta}$ in front of the other nine terms not shown. This term arises in the advection of relative vorticity. The $\epsilon_T \epsilon_R^2$ arises from the local change in relative vorticity and is exactly,

$$\frac{\epsilon_T \epsilon_R^2 G^2}{\sin^3\theta} \frac{1}{\cos^2\theta} \frac{\partial}{\partial \bar{t}} \left[\eta_{\bar{y}\bar{y}}\eta_{\bar{x}\bar{x}} + \eta_{\bar{x}\bar{y}}^2\right] \tag{43}$$

With these we can can place the bound on our error as

$$O(\epsilon^3) = \max\left(\epsilon_R^3 G \cdot \frac{1}{\sin^4\theta}, \epsilon_T \epsilon_R^2 G^2 \cdot \frac{1}{\sin^3\theta}\right). \tag{44}$$

c. Bounding the Cartesian Approximation

Although we have not yet made any assumptions that restrict equation (42) to a Cartesian plane, it is often desirable to make this assumption because the equations become much simpler to solve. A trigonometric factor such as $\sin\theta$ is defined as $\sin\left(\theta_0 + \frac{\epsilon_\beta}{\epsilon_R}\bar{y}\right)$, using equations (30) and (36), and a Taylor expansion about $\theta = \theta_0$ approximates this as $\sin\theta \approx \sin\theta_0 + \cos\theta_0 \frac{\epsilon_\beta}{\epsilon_R}\bar{y}$. Therefore, having already established that $\epsilon_\beta \ll \epsilon_R$ for the scales that we're interested in, the leading order terms to the Cartesian approximation of equation (42) are found by replacing $\sin\theta = \sin\theta_0$, $\cos\theta = \cos\theta_0$ and $\overline{\cos\theta} = 1$.

The next to leading order terms in the Cartesian approximation will contain factors of $\frac{\epsilon_\beta}{\epsilon_R}\bar{y}$ multiplying each term in equation (42). If we want to avoid explicit reference to $\bar{y}$ when we approximate the equation, we will need to estimate the size of these next to leading order terms. However, rather than do this for all the terms, it will later be shown that the ϵ_β, ϵ_R^2 and $\epsilon_H\epsilon_T$ terms generally dominated the others for the scales that we're interested in. The Cartesian approximation for these three terms is bounded by,

$$O(\text{spherical}) = \max\left(\frac{\epsilon_\beta^2}{\epsilon_R} G \cdot \cot^2\theta_0, \epsilon_R\epsilon_\beta G^2 \cdot \frac{3 - 5\cos^2\theta_0}{\sin^3\theta_0\cos\theta_0}, \frac{\epsilon_H\epsilon_T\epsilon_\beta}{\epsilon_R} \cdot \cos\theta_0\right). \tag{45}$$

d. Magnitude of Terms

We can now estimate the magnitude of all the terms shown in equation (42) using their coefficients. The scales found in the observations Chelton et al. (2007) can be inserted into these coefficients directly. For example, take a near geostrophic ($G = 1$) eddy at latitude 45, with height N_0 of 10 cm, length scale L of 100 kilometers and time scale T of 100 days so that L/T is close to the long wave phase speed of a first mode baroclinic Rossby wave with an equivalent depth D of 80 cm. These choices require that the fluid velocity U be at 6.7 cm/s in order to satisfy $G = 1$. The coefficient $\epsilon_\beta G \cot\theta_0$ is approximately $7.2 \cdot 10^{-5}$, whereas $\epsilon_T \epsilon_R G \csc\theta_0$ is an order of magnitude smaller at $5.6 \cdot 10^{-6}$. These are both bigger than the largest $O(\epsilon^3)$ term at $3.9 \cdot 10^{-7}$ and the largest O(spherical) term at $1.1 \cdot 10^{-6}$. Doing this calculation for the all of the remaining terms in equation (42) allows us to estimate which terms are most likely to be important in describing such an eddy.

For fixed latitude, eddy height, geostrophic parameter and ratio L/T, figure (1) shows the magnitude of the terms in equation (42) versus the length scale. Of the 18 $O(\epsilon)$ and $O(\epsilon^2)$ terms in equation (42), only seven appear above the $O(\epsilon^3)$ and O(spherical) terms for these scales. The $O(\epsilon^3)$ region shows that at small length scales (and time scales) the next term in the iteration is needed to keep the equation valid. The O(spherical) terms indicate that as the length scales increase, the cartesian plane approximation breaks down. This particular limiting bound arises from the $\frac{\epsilon_H \epsilon_T \epsilon_\beta}{\epsilon_R} \cdot \cos\theta_0$ in equation (45).

The estimates of the magnitude of terms in equation (42) and figures like (1) can be used to help construct new theories.

7. Equations of Balance Dynamics

We can fix the four parameters ϵ_H, ϵ_T, ϵ_R and ϵ_β relative to each other using integral powers of ϵ. The terms in equation (42) can then be easily ordered in decreasing magnitude. Different choices in the relative magnitudes of the four parameters reflect different choices in the scales being considered and will result in a different ordering of terms. The leading order terms in the resulting equation will be identical to using the asymptotic approach and expanding $\bar{u}$, $\bar{v}$ and η as a perturbation series in ϵ. All higher order terms would differ, but using the slaving approach by leaving η unexpanded results in the same equations as our iterative approach Warn et al. (1995).

It's important to note that choosing a relationship amongst parameters is not a mathematical requirement nor physical requirement, but simply an abstract tool used in various approaches for finding an equation. Once an equation is found we do not require that all the terms have exactly the same magnitude, just as we do not demand that the scales actually obey the precise relationship of the parameters. The f-plane quasi-geostrophic potential vorticity equation (5) when derived using the asymptotic approach Pedlosky (1987) might lead one to believe that all terms will actually be of order ϵ. The reality is that a solution to a given equation will have different magnitudes for each term at a given point, arguably more like the magnitudes in figure (1). The only mathematical requirement is that these terms sum to zero at all points. The reason for fixing the four parameters relative to one another is therefore to provide a concrete ordering amongst terms and in the case of the leading order terms, indicate an equation that can be derived using the asymptotic approach.

a. Quasi-Geostrophic Potential Vorticity, f-plane

The f-plane quasi-geostrophic potential vorticity equation Pedlosky (1987) can be obtained by choosing

$$\epsilon_H = \epsilon, \quad \epsilon_T = \epsilon, \quad \epsilon_R = \epsilon, \quad \epsilon_\beta = \epsilon^3. \tag{46}$$

The leading order terms at $O\left(\epsilon^2\right)$ become the equation of balance dynamics which is just the usual quasi-geostrophic potential vorticity equation on the f-plane,

$$\frac{\epsilon_T\epsilon_R G}{\sin\theta_0}\cdot\nabla^2\eta_{\bar{t}}-\epsilon_H\epsilon_T\sin\theta_0\cdot\eta_{\bar{t}}+\frac{\epsilon_R^2G^2}{\sin^2\theta_0}\cdot\hat{J}\left(\eta,\nabla^2\eta\right)=0. \tag{47}$$

Because all three terms in equation (47) are order ϵ^2, this equation is identical to what would be found using the asymptotic approach as is usually done Pedlosky (1987). With two degrees of freedom in the scales (time and length), we can adjust two of the coefficients to match the third so that,

$$\nabla^2\eta_{\bar{t}}-\eta_{\bar{t}}+\hat{J}\left(\eta,\nabla^2\eta\right)=0 \tag{48}$$

where time is now in units of f_0^{-1} and length in units of the Rossby length scale $L_R=\frac{\sqrt{gD}}{f_0}$. Applying the scales (46) to the shallow water equations (33)-(35) we find,

$$\sin\theta_0\bar{v}=G\frac{\partial\eta}{\partial\bar{x}}+\epsilon_T\frac{\partial\bar{u}}{\partial\bar{t}}+\epsilon_R\left[\bar{u}\frac{\partial\bar{u}}{\partial\bar{x}}+\bar{v}\frac{\partial\bar{u}}{\partial\bar{y}}\right]+O(\epsilon^2), \tag{49}$$

$$\sin\theta_0\bar{u}=-G\frac{\partial\eta}{\partial\bar{y}}-\epsilon_T\frac{\partial\bar{v}}{\partial\bar{t}}-\epsilon_R\left[\bar{u}\frac{\partial\bar{v}}{\partial\bar{x}}+\bar{v}\frac{\partial\bar{v}}{\partial\bar{y}}\right]+O(\epsilon^2) \tag{50}$$

$$0=\frac{\epsilon_H\epsilon_T}{\epsilon_R}\eta_{\bar{t}}+\epsilon_H\left(\bar{u}\frac{\partial\eta}{\partial\bar{x}}+\bar{v}\frac{\partial\eta}{\partial\bar{y}}\right)+(1+\epsilon_H\eta)\left(\frac{\partial\bar{u}}{\partial\bar{x}}+\frac{\partial\bar{v}}{\partial\bar{y}}\right)+O(\epsilon^2). \tag{51}$$

Equations (49)-(51) are appropriate momentum and continuity equations for these scales. The balance relations could be constructed from (49) and (50) by taking the first iteration so that,

$$\bar{v}=G\frac{1}{\sin\theta_0}\frac{\partial\eta}{\partial\bar{x}}-\frac{\epsilon_T G}{\sin\theta_0}\frac{\partial^2\eta}{\partial\bar{t}\partial\bar{y}}+\frac{\epsilon_R G}{\sin\theta_0}\left[\frac{\partial\eta}{\partial\bar{y}}\frac{\partial^2\eta}{\partial\bar{x}\partial\bar{y}}-\frac{\partial\eta}{\partial\bar{x}}\frac{\partial^2\eta}{\partial\bar{y}\partial\bar{y}}\right], \tag{52}$$

$$\bar{u}=-G\frac{1}{\sin\theta_0}\frac{\partial\eta}{\partial\bar{y}}-\frac{\epsilon_T G}{\sin\theta_0}\frac{\partial^2\eta}{\partial\bar{t}\partial\bar{x}}+\frac{\epsilon_R G}{\sin\theta_0}\left[\frac{\partial\eta}{\partial\bar{y}}\frac{\partial^2\eta}{\partial\bar{x}\partial\bar{x}}-\frac{\partial\eta}{\partial\bar{x}}\frac{\partial^2\eta}{\partial\bar{x}\partial\bar{y}}\right]. \tag{53}$$

However, equation (47) is an equation for the height field η only at the leading order and so there may be variations in η of $O(\epsilon)$. Equations (52) and (53) are really only valid to leading order and so we must take

$$\bar{v} = G\frac{1}{\sin\theta_0}\frac{\partial\eta}{\partial\bar{x}}, \tag{54}$$

$$\bar{u} = -G\frac{1}{\sin\theta_0}\frac{\partial\eta}{\partial\bar{y}}, \tag{55}$$

as our balance relations.

b. Quasi-Geostrophic Potential Vorticity, β-plane

The β-plane quasi-geostrophic potential vorticity equation Pedlosky (1987) can be obtained by choosing

$$\epsilon_H = \epsilon, \quad \epsilon_T = \epsilon, \quad \epsilon_R = \epsilon, \quad \epsilon_\beta = \epsilon^2. \tag{56}$$

All terms are again leading order terms at $O\left(\epsilon^2\right)$ so that equation (42) becomes

$$\frac{\epsilon_T\epsilon_R G}{\sin\theta_0}\cdot\nabla^2\eta_{\bar{t}} - \epsilon_H\epsilon_T\sin\theta_0\cdot\eta_{\bar{t}} + \epsilon_\beta G\cot\theta_0\cdot\eta_{\bar{x}} + \frac{\epsilon_R^2 G^2}{\sin^2\theta_0}\cdot\hat{J}\left(\eta,\nabla^2\eta\right) = 0 \tag{57}$$

With two fundamental units and four coefficients in equation (57), we cannot completely eliminate all the coefficients. Using the same units as with the f-plane, this places a non-dimensional coefficient typically denoted as β in front $\eta_{\bar{x}}$ term. However, the time scale of f_0^{-1} is unnaturally short for mesoscale features. If one takes the time scale as $\beta_0^{-1}L_R^{-1}$ then equation (57) can be scaled to

$$\nabla^2\eta_{\bar{t}} - \eta_{\bar{t}} + \eta_{\bar{x}} + \beta^{-1}\cdot\hat{J}\left(\eta,\nabla^2\eta\right) = 0 \tag{58}$$

where $\beta^{-1} = \frac{\sqrt{gD}}{\beta_0 L_R^2}$. The non-dimensionalized momentum equations now contain variations of the Coriolis parameter $\bar{f} = \sin\theta_0 + \frac{\epsilon_\beta}{\epsilon_R}\cos\theta_0 \bar{y}$ and a metric term $\gamma = 1 - \frac{\epsilon_\beta}{\epsilon_R}\tan_{\theta_0}\bar{y}$,

$$\bar{f}\bar{v} = \frac{G}{\gamma}\frac{\partial\eta}{\partial\bar{x}} + \epsilon_T\frac{\partial\bar{u}}{\partial\bar{t}} + \epsilon_R\left[\bar{u}\frac{\partial\bar{u}}{\partial\bar{x}} + \bar{v}\frac{\partial\bar{u}}{\partial\bar{y}}\right] + O(\epsilon^2), \tag{59}$$

$$\bar{f}\bar{u} = -G\frac{\partial\eta}{\partial\bar{y}} - \epsilon_T\frac{\partial\bar{v}}{\partial\bar{t}} - \epsilon_R\left[\bar{u}\frac{\partial\bar{v}}{\partial\bar{x}} + \bar{v}\frac{\partial\bar{v}}{\partial\bar{y}}\right] + O(\epsilon^2), \tag{60}$$

$$\frac{\epsilon_H\epsilon_T}{\epsilon_R}\eta_{\bar{t}} + \epsilon_H\left(\frac{\bar{u}}{\gamma}\frac{\partial\eta}{\partial\bar{x}} + \bar{v}\frac{\partial\eta}{\partial\bar{y}}\right) = -\frac{(1+\epsilon_H\eta)}{\gamma}\left(\frac{\partial\bar{u}}{\partial\bar{x}} + \frac{\partial(\bar{v}\gamma)}{\partial\bar{y}}\right) + O(\epsilon^2). \tag{61}$$

The metric term γ arises as the Taylor expansion of $\overline{\cos}\theta$ and appears to be occasionally neglected Vallis (2006), perhaps because it does not appear in the potential vorticity equation (57) (see Pedlosky (1987) and Ripa (1997) for details). The balance relations remain the same as the f-plane approximation because equation (57) again only contains leading order terms,

$$\bar{v} = G\frac{1}{\sin\theta_0}\frac{\partial\eta}{\partial\bar{x}}, \tag{62}$$

$$\bar{u} = -G\frac{1}{\sin\theta_0}\frac{\partial\eta}{\partial\bar{y}}. \tag{63}$$

c. *Flierl-Petviashvili Equation*

The FP equation arises in several different contexts including the Great Red Spot on Jupiter Petviashvili (1980), as well as modifications to quasi-geostrophy with an exterior mean shear flow Flierl (1979), finite amplitude height changes Anderson and Killworth (1979) Charney and Flierl (1981) and thermobaricity of the equation of state de Szoeke (2004). In our case it can be found with the choice of scales

$$\epsilon_H = \epsilon, \quad \epsilon_T = \epsilon^2, \quad \epsilon_R = \epsilon^2, \quad \epsilon_\beta = \epsilon^3. \tag{64}$$

Unlike the two quasi-geostrophic potential vorticity equations (47) and (57), the FP equation requires both the leading order and next to leading order terms from equation (42). This means that it is not possible to find this equation using the asymptotic approach and letting $\epsilon \mapsto 0$. Retaining the $O(\epsilon^3)$ terms and the $O(\epsilon^4)$ terms we have,

$$\frac{\epsilon_T \epsilon_R G}{\sin\theta_0} \cdot \nabla^2 \eta_{\bar{t}} - \epsilon_H \epsilon_T \sin\theta_0 \cdot \eta_{\bar{t}} (1 + \epsilon_H \eta)^{-1} + \epsilon_\beta G \cot\theta_0 \cdot \eta_{\bar{x}} + \frac{\epsilon_R^2 G^2}{\sin^2\theta_0} \cdot \hat{J}\left(\eta, \nabla^2 \eta\right) = 0. \quad (65)$$

Using the same time and length scale as with equation (57), η can be rescaled by a factor of ϵ_H^{-1}, with the additional requirement that $\eta \ll 1$. In this case the FP equation becomes,

$$\eta_{\bar{t}} - \eta_{\bar{t}}(1 - \eta) + \eta_{\bar{x}} + \beta^{-1} \cdot \hat{J}\left(\eta, \nabla^2 \eta\right) = 0, \quad (66)$$

where we've used the fact the η is small to write $(1+\eta)^{-1}$ as $(1-\eta)$. The $\eta_{\bar{t}}$ and $\eta_{\bar{x}}$ terms are $O(\epsilon^3)$, while the remaining terms are $O(\epsilon^4)$. Therefore we can multiply the equation by $(1+\eta)$ also consistently write,

$$\eta_{\bar{t}} - \eta_{\bar{t}} + \eta_{\bar{x}}(1 + \eta) + \beta^{-1} \cdot \hat{J}\left(\eta, \nabla^2 \eta\right) = 0. \quad (67)$$

The balance relations will now also contain the next to leading order terms of the momentum equations (with factors of $\frac{\epsilon_\beta}{\epsilon_R} = \epsilon$),

$$\bar{v} = G \frac{1}{\bar{f}\gamma} \frac{\partial \eta}{\partial \bar{x}}, \quad (68)$$

$$\bar{u} = -G \frac{1}{\bar{f}} \frac{\partial \eta}{\partial \bar{y}}. \quad (69)$$

d. FP + J″ Equation

Finally consider the choice of scales,

$$\epsilon_H = \epsilon, \quad \epsilon_T = \epsilon^3, \quad \epsilon_R = \epsilon^2, \quad \epsilon_\beta = \epsilon^4. \quad (70)$$

Using the same choice of time and length scale as the FP equation, we have

$$\eta_{\bar{t}} - \eta_{\bar{t}}(1-\eta) + \eta_{\bar{x}} + \beta^{-1} \cdot \hat{J}\left(\eta, \nabla^2\eta\right) - \beta^{-1} \cdot J'' = 0, \tag{71}$$

where $J'' = \frac{1}{\sin^2\theta}\left[\eta_{\bar{x}}\eta_{\bar{y}}\left(\eta_{\bar{x}\bar{x}} - \eta_{\bar{y}\bar{y}}\right) + \eta_{\bar{x}\bar{y}}\left(\eta_{\bar{y}}^2 - \eta_{\bar{x}}^2\right)\right]$ following the convention of Williams and Yamagata (1984). This term arises as the advection of η, an ageostrophic effect. To see the origin of this term, take the expression for the advection of η,

$$\epsilon_R\left(\frac{\bar{u}}{\cos\theta}\frac{\partial\eta}{\partial\bar{x}} + \bar{v}\frac{\partial\eta}{\partial\bar{y}}\right)$$

and substitute in equations (33) and (34), ignoring all but the ϵ_R components. The resulting equation will still contain $\bar{u}$ and $\bar{v}$ that can then be replaced with their geostrophic approximations.

The balance relations do not contain the geometric effects found in the FP balance relations (68) and (69) because the factor $\frac{\epsilon_\beta}{\epsilon_R}$ is $O(\epsilon^2)$. The balance relations are thus,

$$\bar{v} = G\frac{1}{\sin\theta_0}\frac{\partial\eta}{\partial\bar{x}}, \tag{72}$$

$$\bar{u} = -G\frac{1}{\sin\theta_0}\frac{\partial\eta}{\partial\bar{y}}. \tag{73}$$

8. Valid Regimes

Using the information from figure (1) we can estimate what terms a theory would need in describing these eddies and evaluate the a priori appropriateness of an existing theory. For example, the five terms with the largest magnitude at 150 km in figure (1) are exactly the same terms found in the FP equation (66). From about 60 km to 140 km, the terms from equation (71) dominate. Nowhere do the terms from quasi-geostrophy (58) dominate the others. This would suggested that, at least for latitude 45, an extended theory like equation (71) is necessary for capturing the appropriate physics of the eddies.

Figure (2) shows the regions where different equations are valid, in the sense that the terms of those equations dominate all other neglected terms, as a function of length scale and latitude. Regions of blue indicate areas where none of the theories were valid without adding additional, neglected terms. From figure (2) it is clear that the quasi-geostrophic potential vorticity equation (58) is only valid for a band of length scales in the small region around 15 degrees of latitude. Between latitudes 15 degrees and 60 degrees various extensions to QG theory are required with additional terms, all arising from a relatively large surface height to equivalent depth ratio, ϵ_H.

More abstractly, each of the regions shown in figure (2) are actually five-dimensional subregions inside the full five-dimensional parameter space defined by the coordinates (θ_0, N_0, U, L, T). Figure (2) is only showing a particular two-dimensional slice through the parameter space. The values we assign to the scales are only estimates, and so a theory occupying a larger subregion around our estimated values is more robust. Figures (3) and (4) show nearby regions in the parameter space by varying the L/T ratio and N_0, respectively. These suggest the extend QG theories occupy a fairly larger region of parameter space around the scales of interest.

The two equations most likely to describe the observations at mid-latitudes are the FP equation (66) at length scales roughly 125-175 km and a new extension to the FP equation with the J$''$ term (71) for length scales roughly 60 - 125 km. Although the FP equation is a subset of terms from (71), we find that it is generally not true that the FP equation alone is valid for the same regions as (71) (the J$''$ term is not the smallest term). It also worth noting that the length scale hierarchy of the two equations is reflected in the scaling and balance relations found in the previous section. In particular, the balance relations (68) and (69) include a geometric and β-plane effect not found in the balance relations (72) and (73), consistent with the notion that the FP equation is valid at longer length scales ((66).

9. Beyond Geostrophy

In the previous section we considered only flows where the scales were chosen to reflect the geostrophic balance, $G = 1$. The reality, of course, is that we expect the flows to deviate from pure geostrophy and hope that our equation remains valid. Because we left the geostrophic parameter unfixed in our derivation of equation (42) we can consider flows that depart from $G = 1$ and determine when our equation remains valid.

Figure (5) shows that our theories are valid for a wide range of fluid speeds deviating from a pure geostrophic balance, from $G = 10^{-1}$ to $G = 10^{5}$. In fact, it is interesting to note that when considering flows that depart from geostrophy, suitable equations can still be found that accurately describe the flows being considered.

10. The FP Monopole

Once an equation has been selected as an appropriate description of some physical phenomenon, the next task is typically to find solutions admitted by the chosen equation. However, not all solutions to the given equation are necessarily valid; they may violate assumptions made during the equation's derivation. As an extreme example consider a plane wave of height 100 m and length scale 1 mm as a solution to equation (48). Although it true that such a plane wave indeed solves equation (48), it violates a number of assumptions including the small amplitude assumption and neglecting of viscosity. We can use the same techniques used above to show the validity of a given equation to assess the validity of a particular solution to an equation.

The monopole solution to equation (67) Petviashvili (1980); Flierl (1979) is anticyclonic, axially symmetry and propagates slightly faster than the long wave Rossby wave phase speed. This solution is therefore naturally of interest as it may represent some of the satellite altimetry tracked eddies observed by Chelton et al. (2007). The solution can be written in approximated form Boyd (1991) as

$$\eta_{\mathrm{FP}}\left(r = \frac{\sqrt{x^2+y^2}}{L}\right) = \frac{2gh^2}{L^2 f_0^2 \alpha} \cdot \frac{2.3822 + 1.01327r^2 - 0.02417r^4}{1 + 0.75r^2 + (1/16)r^4 + (1/64)r^6}. \tag{74}$$

The parameter α is determined from the vertical structure and is typically 1.3 Flierl (1979). The FP monopole solution (74) has one free parameter, L, which sets the length scale of the monopole while simultaneously setting the height of the monopole. If we take the height of the monopole to be 10 centimeters at latitude 24, then we find that the e-folding diameter of the monopole is 725 km! This unreasonably large and almost certainly violates a number of assumptions. On the other hand, the same monopole of height 10 cm at latitude 55 only has an e-folding diameter of 350 km.

To check the FP monopole solution against our assumptions rigorously, we assume that if the FP monopole (74) is a valid solution to equation (67), then all five terms in (67) must be of greater magnitude than all the neglected terms in equation (42). The yellow region in figure (6) shows that the FP monopole solution (74) is therefore valid for only a vary narrow range of length scales at at high latitudes. The length scales in figure (6) were taken to be an approximate e-fold radius of the FP monopole, suggesting that only monopoles of diameter greater than 320 km are valid. Based on the observations in Chelton et al. (2007), this suggests that while the FP equation itself might be valid (as indicated in figures (3) and (4)), the FP monopole itself does not likely represent any sizable number of the observed eddies.

11. Conclusions

By developing a new approach that retains freedom in the scales of motion, we were able to find that a new extension to the FP equation with the J'' term (71) is evidently required to describe the mid-latitude mesoscale eddies observed by Chelton et al. (2007). Eddies with length scales above 125 km are well described by the FP equation (66), but additional corrections to the Cartesian approximation are required for lengths beyond 175 km. Both of these theories extend quasi-geostrophic theory by including effects attributed to the large surface height displacement relative to the equivalent depth. This directly pinpoints the small-amplitude assumption as the primary weakness in quasi-geostrophic theory at these scales.

This technique provides a concrete way of assessing the range of scales over which a given theory is valid. Although not a rigorous proof a theory's validity, this does provide an estimate of when caution needs to be exercised for a theory's application toward a particular problem. From these arguments it would appear that equation (71) range of valid scales dominates the regions of interest. Whether this extensions to quasi-geostrophic theory will show important differences in the evolution of eddies in a model remains to be shown.

The scales considered here were chosen to reflect the satellite altimetry observations, but a wide range of other choices could certainly be considered. We expect that quasi-geostrophic theory will dominate in a large region of parameter space around other values, such as smaller height perturbations and regions where the length scale is not fixed to the time scale. Similarly, different choices in length scale, time scale, surface height and latitudes may all yield very different equations suitable for different contexts. Finally, the lowest order expansions in the cartesian approximation could be used to consider theories containing weak dependence on latitude.

Acknowledgments.

This research was supported by the National Science Foundation, Award 0621134, and by the National Aeronautics and Space Administration, grant NNX08AR37G.

REFERENCES

Ablowitz, M. and P. Clarkson, 1991: *Solitons, nonlinear evolution equations and inverse scattering*. Cambridge University Press.

Allen, J., 1993: Iterated geostrophic intermediate models. *Journal of Physical Oceanography*, **23 (11)**, 2447–2461.

Anderson, D. and P. Killworth, 1979: Nonlinear propagation of long rossby waves. *Deep-Sea Res*, **26**, 1033–1049.

Bluman, G. and S. Anco, 2002: *Symmetry and integration methods for differential equations*. Springer.

Boyd, J., 1991: Monopolar and dipolar vortex solitons in two space dimensions. *Wave Motion*, **13 (3)**, 223–241.

Charney, J. and G. Flierl, 1981: Oceanic analogues of large-scale atmospheric motions. *Evolution of Physical Oceanography*, 504–548.

Chelton, D. and M. Schlax, 1996: Global observations of oceanic rossby waves. *Science*, **272 (5259)**, 234.

Chelton, D. B., M. G. Schlax, R. M. Samelson, and R. A. de Szoeke, 2007: Global observations of large oceanic eddies. *Geophys. Res. Lett.*, **34**, URL `http://dx.doi.org/10.1029/2007GL030812`.

de Szoeke, R., 2004: An effect of the thermobaric nonlinearity of the equation of state: A mechanism for sustaining solitary rossby waves. *J. Phys. Oceanogr.*, **34 (9)**, 2042–2056.

Flierl, G., 1979: Baroclinic solitary waves with radial symmetry. *Dyn. Atmos. Oceans*, **3**, 15–38.

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

Gill, A., 1982: *Atmosphere-ocean dynamics*. Academic press.

Killworth, P., D. Chelton, and R. de Szoeke, 1997: The speed of observed and theoretical long extratropical planetary waves. *J. Phys. Oceanogr.*, **27 (9)**, 1946–1966.

Lighthill, M., 1952: On sound generated aerodynamically. i. general theory. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 564–587.

Pedlosky, J., 1987: *Geophysical Fluid Dynamics*. 2d ed., Spring-Verlag.

Petviashvili, V., 1980: Red spot of jupiter and the drift soliton in a plasma. *Journal of Experimental and Theoretical Physics Letters*, **32**.

Ripa, P., 1997: "inertial" oscillations and the β-plane approximation(s). *J. Phys. Oceanogr.*, **27**, 633–647.

Vallis, G., 1996: Potential vorticity inversion and balanced equations of motion for rotating and stratified flows. *Quart. J. Roy. Meteor. Soc.*, **122 (529)**.

Vallis, G., 2006: *Atmospheric and oceanic fluid dynamics: fundamentals and large-scale circulation*. Cambridge University Press.

Warn, T., O. Bokhove, T. Shepherd, and G. Vallis, 1995: Rossby number expansions, slaving principles, and balance dynamics. *Quart. J. Roy. Meteor. Soc.*, **121 (523)**.

Williams, G. and T. Yamagata, 1984: Geostrophic regimes, intermediate solitary vortices and jovian eddies. *J. Atmos. Sci.*, **41 (4)**, 453–478.

List of Figures

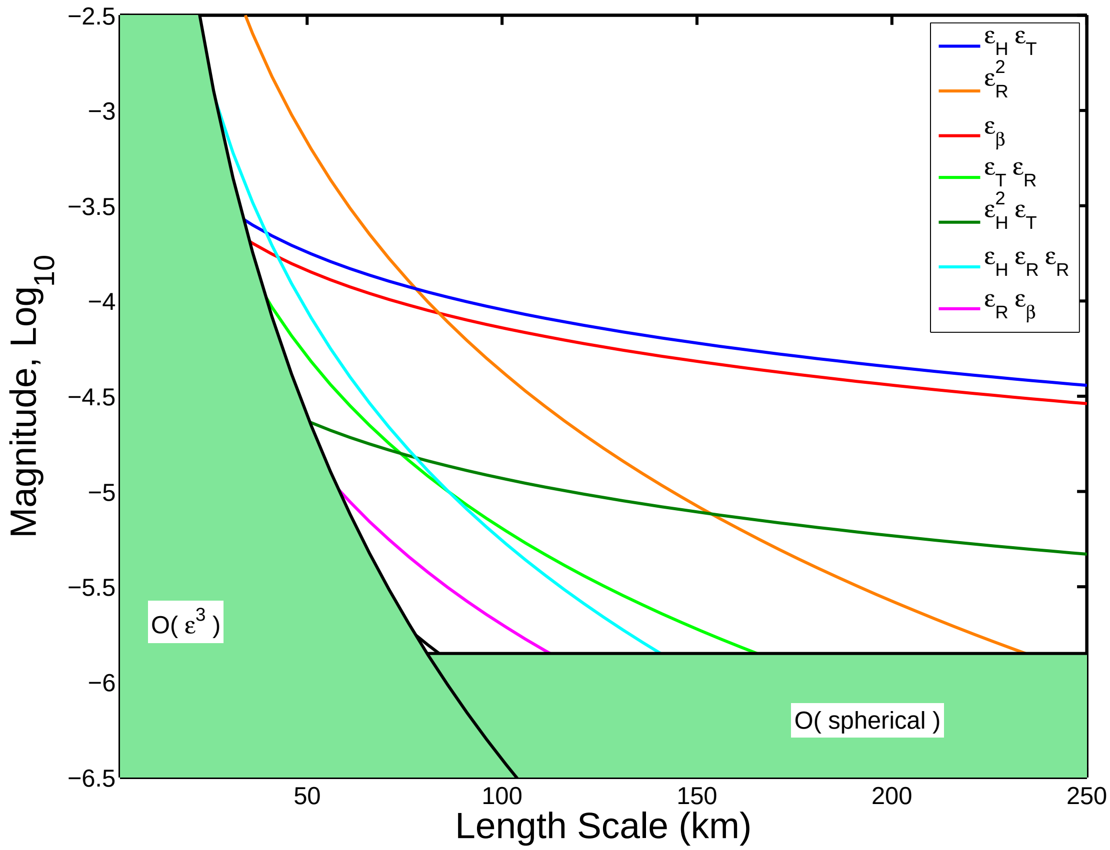

FIG. 1. The magnitude of the terms in equation (42) are shown versus the length scale. The coefficients in the legend can be matched with the coefficients in equation (42). The surface height is fixed at 10 cm, the latitude at 45 and the fluid velocity varies assuming $G = 1$. The ratio of the length scale to the time scale is fixed at 1.46 cm/s, slightly faster than the long wave Rossby phase speed at that latitude.

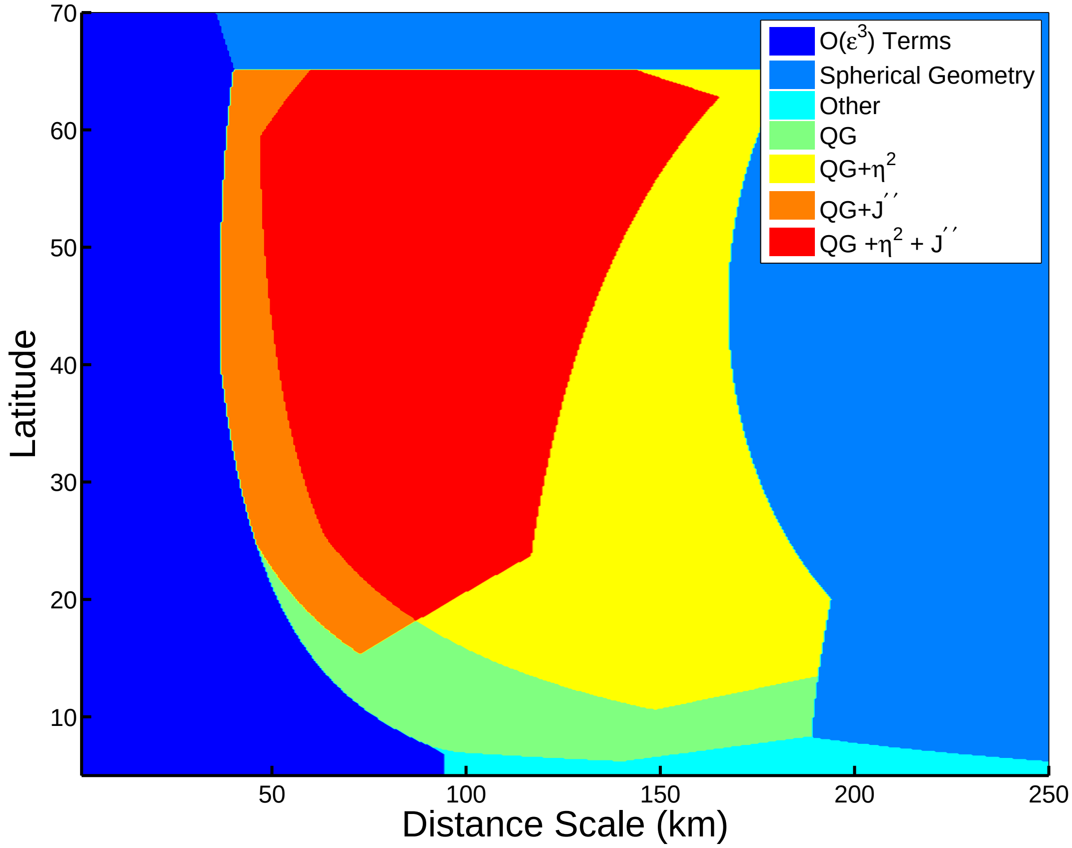

FIG. 2. Each colored region shows the area in parameter space where a given equation is valid. An equation is declared valid in a given region if all of its terms have the greatest magnitude. This is a 'best case scenario' for a equation as one might typically demand that its terms be a factor of 2 or 5 larger than other neglected terms and therefore cause the regions to shrink. The blue regions require additional terms from either the next order expansion $O(\epsilon^3)$, or from expansion of trigonometric terms $O(\text{spherical})$. The surface height is fixed at 10 cm ($\epsilon_H = 0.13$), $G = 1$ and the ratio of the length scale to the time scale is fixed at the long Rossby wave phase speed. The equations analyzed are the quasi-geostrophic potential vorticity equation (58) (QG), the FP equation (66) (QG + η^2), equation (71) (QG + η^2 + J$''$) and a fourth equation extending QG with the J$''$ term (QG + J$''$).

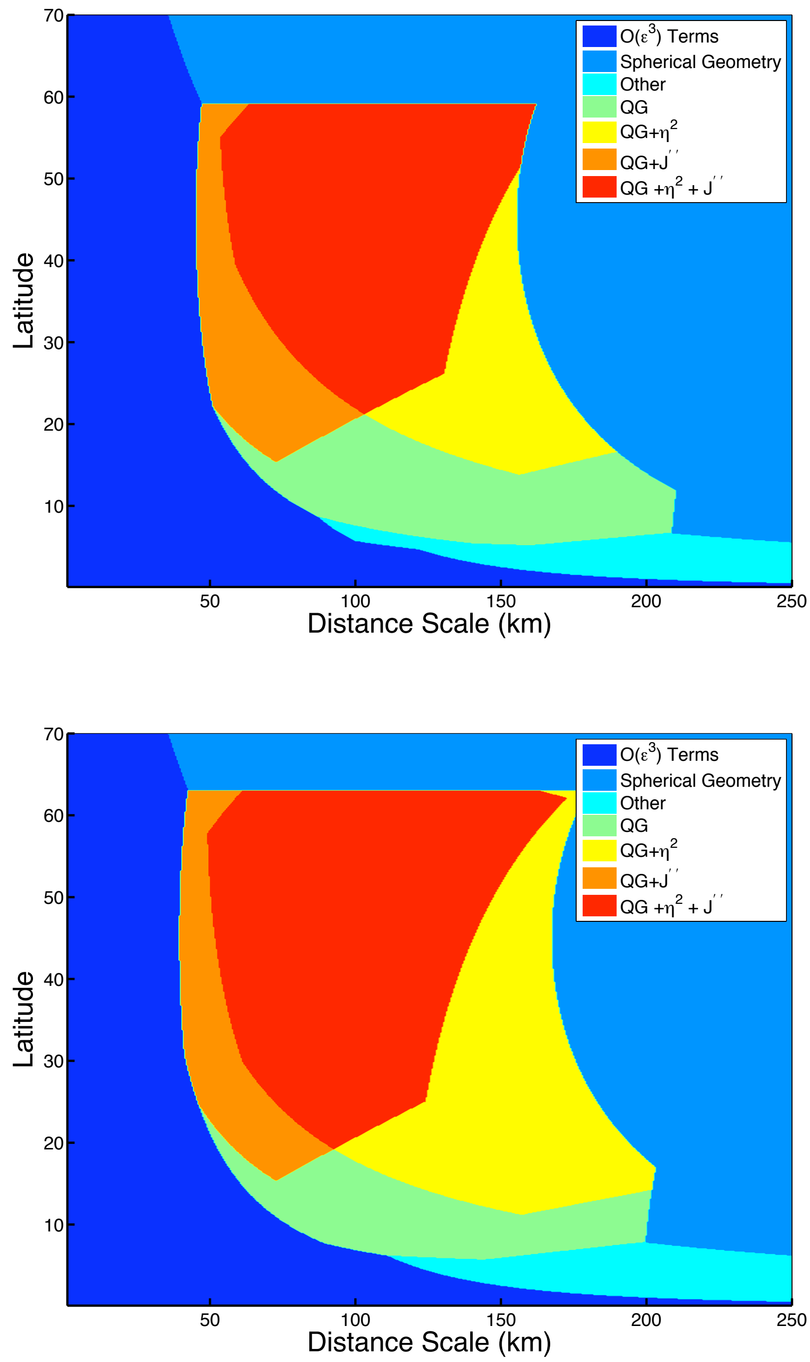

FIG. 3. Similar to figure (2), but now the ratio of the length scale to the time scale is fixed at 0.8 x the long Rossby wave phase speed in the first panel and 1.2 x the phase speed in the second panel. This suggests that the extended QG theories are valid over a range of phase speeds.

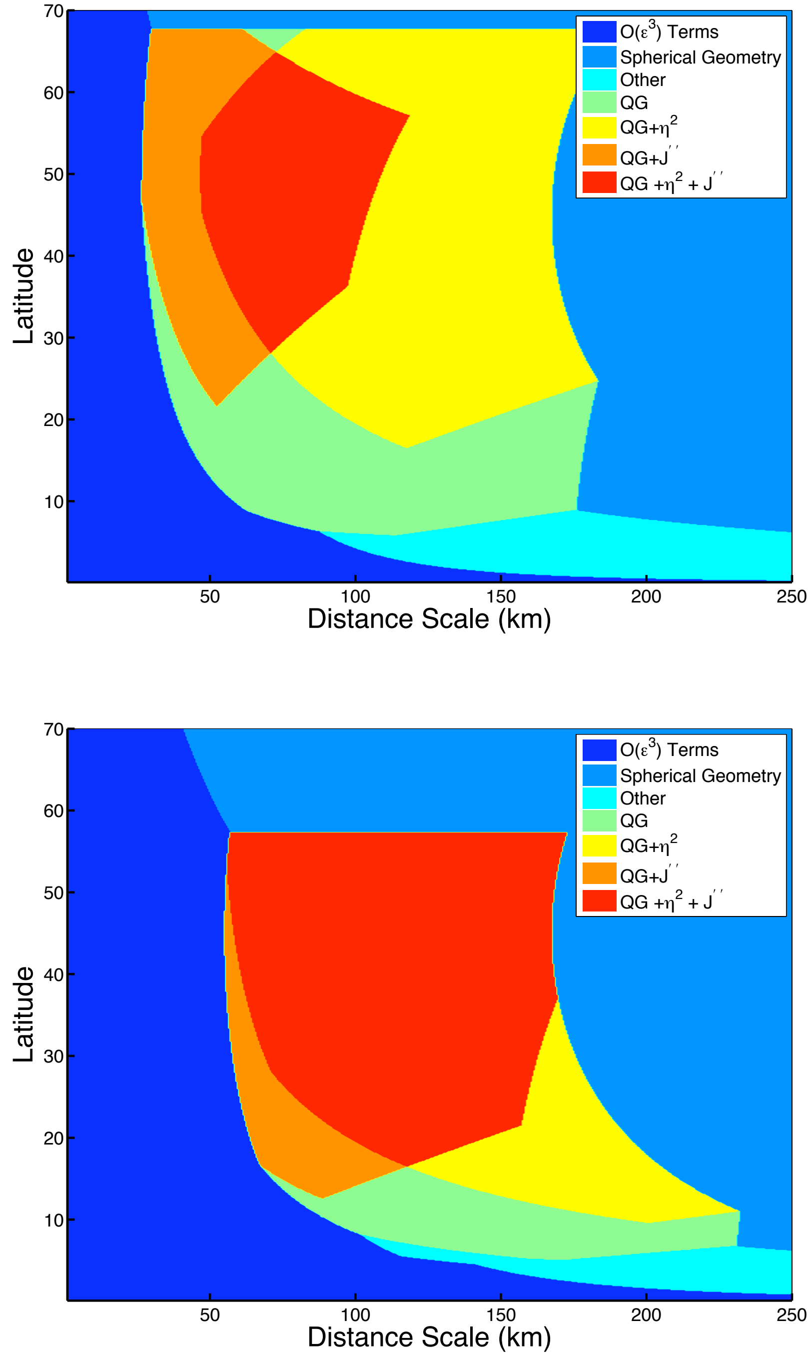

FIG. 4. Similar to figure (2), but now the surface height is fixed at 5 cm ($\epsilon_H = 0.63$) in the first panel and 15 cm ($\epsilon_H = 0.19$) in the second panel. Quasi-geostrophic theory performs better at the smaller height perturbations, whereas equation (71) (QG + η^2 + J$''$ dominates at the larger perturbations.

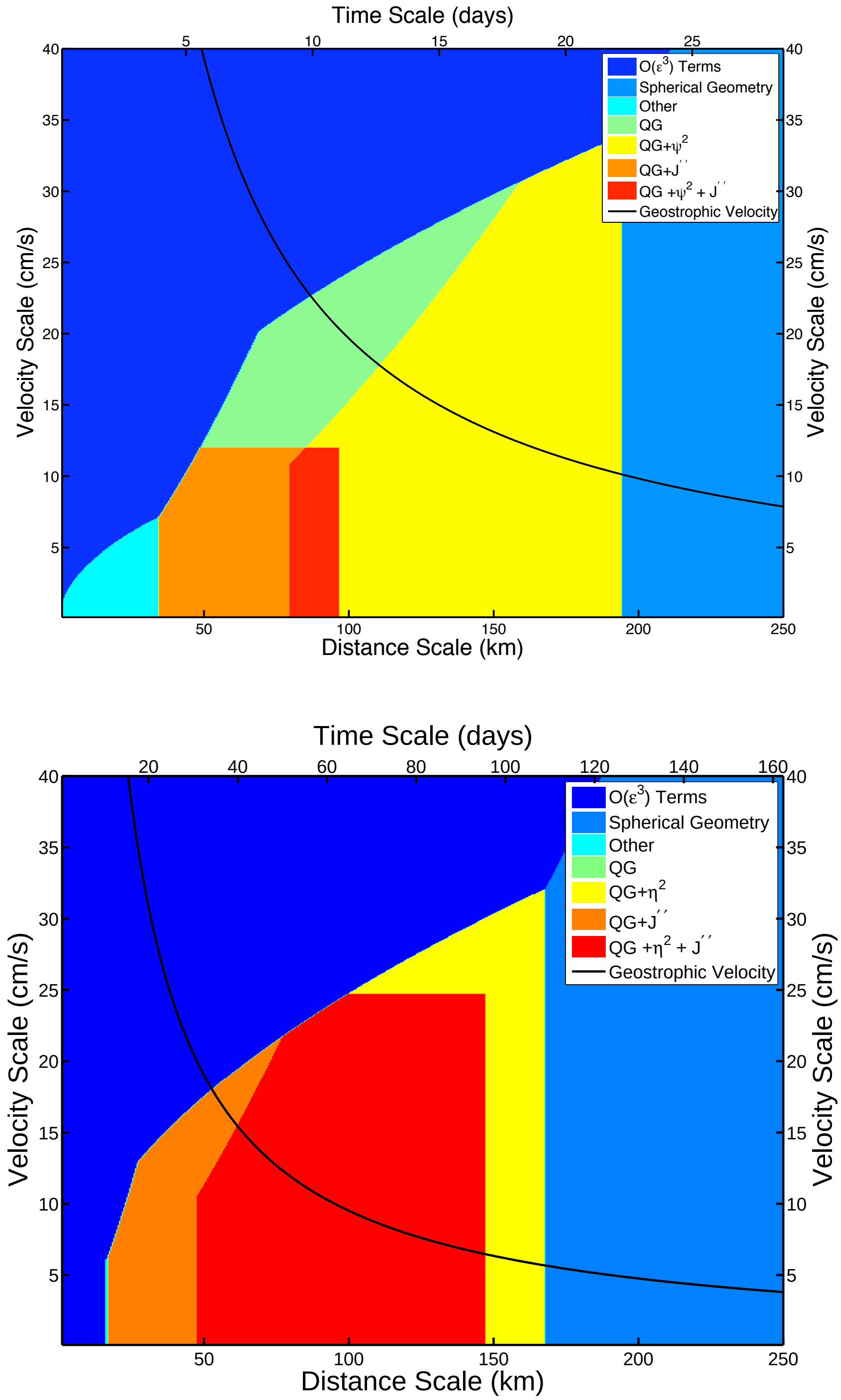

FIG. 5. Similar to figure (2), but now the latitude is fixed and the velocity is allowed to vary (allowing departures from $G = 1$). The black line is $G = 1$. The surface height is fixed at 10 cm ($\epsilon_H = 0.13$) and the ratio of the length scale to the time scale is fixed at 1.5 x the long Rossby wave phase speed for the top panel at 20 degrees, and 1.0 x the Rossby wave phase speed for the bottom panel at latitude 45.

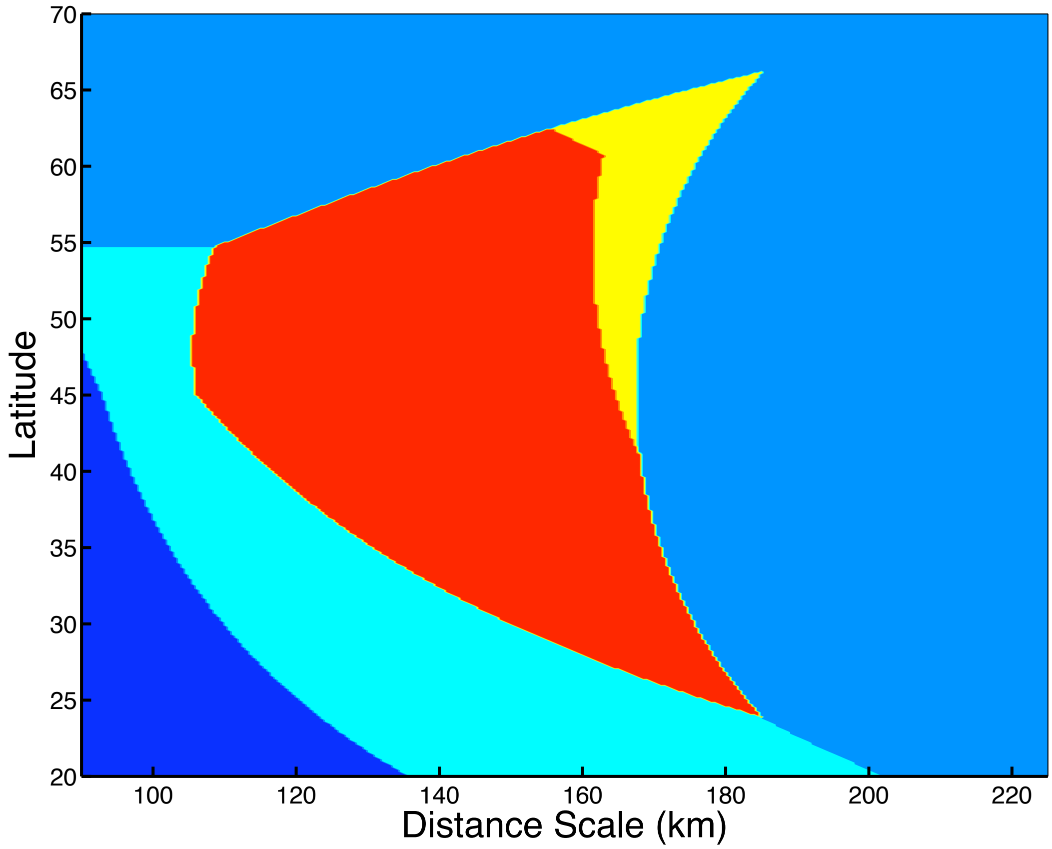

FIG. 6. Similar to figure (2), but now the scales are fixed for the FP monopole solution (74). The yellow region indicates the small region in parameter space where the five terms from the FP equation (67) dominate all other neglected terms. By assumption then, this is the only region where the FP monopole solution (74) is valid. The distance scale shown here is taken to be the e-fold radius of the FP solution.