1. INTRODUCTION

Balance is a central theme in geophysical fluid dynamics. The atmosphere displays vortical and inertia-gravity wave motion. However, observations suggest that some of this motion may be considered as noise. A balanced theory of atmospheric motion would propose relationships between the dynamic variables that eliminate or minimize this noise, thus providing a clearer picture of the physics.

Classically, there are two approaches corresponding to two distinct balance regimes. The small Rossby number limit (Charney, 1948) corresponds to a rapidly rotating flow and was originally developed for the tropics, where the Rossby equations. The small Froude number limit (Charney, 1963) corresponds to a strongly stratified flow and was originally developed for the tropics, where the quasi-geostrophic theory would incorporate both cases systematically and simultaneously. It is hoped that such an approach might ultimately yield uniformly valid balance equations over the entire sphere. This would solve a failure of geostrophic balance since these equations break down at the equator where the Rossby number tends to infinity.

Fundamentally, balance is predicted when the dynamics admits motion with vastly different time scales. For geophysical flows, balance requires the time scale of the vortical motion to be significantly larger than that of the inertia-gravity waves. In the language of Warn et al. (1995), the vortical motion is described by slow variable, \( s \), and the inertia-gravity waves by fast variables, \( f \). The balance theory is complete when we can write the equations in the form

\[
\begin{align*}
\frac{\partial s}{\partial t} &= S(s, f; \epsilon), \\
\frac{\partial f}{\partial t} + \Lambda f &= F(s, f; \epsilon),
\end{align*}
\]

where \( \Lambda \) is a linear invertible operator, \( S \) and \( F \) are nonlinear functions, and \( \epsilon \) is a small asymptotic parameter. We shall therefore define \( \epsilon \) to be a dimensionless time scale separation parameter, small values of which will characterize balance, rather than a particular balance regime. This simple fact makes \( \epsilon \) the correct asymptotic parameter for a general theory of balance. No further assumptions are necessary. In particular, the balance of terms in the momentum equations is not considered.

2. THE \( f \)-PLANE EQUATIONS

First, consider the rotating shallow-water equations in standard Cartesian coordinates,

\[
\begin{align*}
\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} - f v + g \frac{\partial h}{\partial x} &= 0, \\
\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + f u + g \frac{\partial h}{\partial y} &= 0, \\
\frac{\partial h}{\partial t} + v \frac{\partial h}{\partial x} + u \frac{\partial h}{\partial y} + h \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) &= 0,
\end{align*}
\]

where \((u, v)\) is the horizontal fluid velocity, \( f \) is the (constant) Coriolis parameter, \( h \) is surface height and \( g \) is the constant acceleration due to gravity.

Linearizing about a rest state and assuming a non-linear frequency broadening, the dispersion relation has solutions

\[
\omega_0 \sim U\kappa, \quad \omega_{\pm} \sim U\kappa \pm \sqrt{f^2 + gH\kappa^2},
\]

where \( U \) is some characteristic flow speed, \( H \) is the mean fluid depth, \( \kappa^2 = k^2 + \ell^2 \) and \( k \) and \( \ell \) are wave numbers in the \( x \) and \( y \) directions respectively. The ratio of the fast time scale to the slow time scale is

\[
\frac{\omega_0}{\omega_{\pm}} \sim \frac{1}{1 + \frac{\kappa^2 + \ell^2}{RF}}
\]

where \( R = U\kappa/f \) is the Rossby number and \( F = U/\sqrt{gH} \) is the Froude number. Clearly, this ratio will be small if and only if

\[
\epsilon = \frac{RF}{\sqrt{R^2 + F^2}} \ll 1.
\]

Notice that

\[
\frac{\min[R, F]}{\sqrt{2}} \leq \epsilon \leq \min[R, F].
\]

As an asymptotic parameter for balance, \( \epsilon \) is equivalent to \( \min[R, F] \), and yet was determined solely from the
physics of the equations. We can understand this by realising that a dynamical system, initialized on a slow manifold, will remain there if no fast motion is excited. Looking at the dispersion relation graph (figure 1), this occurs if the potential for frequency matching is minimal. There are, evidently, two ways to do this: either the slope of the vortical motion curve can be made smaller relative to the slope of the inertia-gravity wave curve, i.e. \( F \) can be made small, or the inertia-gravity wave curve can be shifted upwards by an increase in \( V \), i.e. \( R \) can be made small. As noted above, both of these possibilities are captured in the small \( \epsilon \), time scale separation, limit.

Appropriately non-dimensionalising and using new dynamic variables, (3)-(5) take the form (1)-(2):

\[
\frac{\partial q}{\partial t} = -\nabla \cdot \nabla q - J(q, q),
\]

\[
\frac{\partial D}{\partial t} - \frac{\Gamma}{\epsilon} = -J(\phi, \nabla^2 \psi) + 2J \left( \frac{\partial \psi}{\partial x} \frac{\partial \psi}{\partial y} \right) - \frac{1}{2} \nabla^2 \left[ \phi \cdot \nabla \phi + 2J(\psi, \phi) \right],
\]

\[
\frac{\partial V}{\partial t} + \frac{A D}{\epsilon} = (1 - b^2) \nabla^2 \phi \cdot (\eta \nabla \phi) + J(\psi, \phi) + b \left[ \nabla \cdot (\nabla^2 \psi \nabla \phi + J(\psi, \nabla^2 \psi) \right],
\]

where \( q = (\nabla^2 \psi - b\eta)/(1 + \epsilon \eta) \), \( D = \nabla^2 \phi \), \( \Gamma = (b\nabla^2 \psi - (1 - b^2)\nabla^2 \eta) \), \( b = B/\sqrt{1 + B^2} = F/\sqrt{R^2 + F^2} \), \( B = F/R \), \( J(\phi, g) = \frac{\partial \phi}{\partial t} g - \frac{\partial g}{\partial t} \phi \), \( A = b^2 - (1 - b^2)\nabla^2 \) and we have used \( h = 1 + \epsilon \eta \) and \( u, v = \nabla \phi + k \times \nabla \psi \).

There is no singularity in these equations as \( b \) ranges between 0 and 1. Thus \( \epsilon \) is our sole singular perturbation parameter; special limits can subsequently be considered by treating \( b \) (i.e. \( B \)) as a regular perturbation parameter.

The analysis proceeds as is typical of an asymptotic expansion except that we assume the slow variable exact and expand only the fast variables (see Warn et al. 1995). To first order we get

\[
\frac{\partial q}{\partial t} = -J(q_0, q),
\]

\[
= -J \left( \left[ \nabla^2 - \frac{b^2}{1 - b^2} \right]^{-1} q, q \right). \tag{14}
\]

This looks just like the quasi-geostrophic (QG) equation for shallow water flow. However, our derivation shows that it is not a QG equation; no assumption of \( R \ll 1 \) has been made, only that \( \epsilon \ll 1 \). In particular, the flow need not be geostrophic. In the standard QG limit \( R \to 0, F \to 0 \) with \( b \) (i.e. \( B \)) held fixed. Of course, our system (14) applies to this case and is consistent. Higher order models can be obtained systematically. Furthermore, special limits \( R \) fixed, \( F \to 0 \) (small \( F \) limit) and \( F \) fixed, \( R \to 0 \) (PG limit) may also be found, taking care to interpret \( b \) correctly.

The above analysis is limited to a planar geometry. Before we can consider the sphere we must understand the physical differences between the two. Significantly, the Rossby number, \( U_k/f \), tends to infinity at the equator; clearly small Rossby number (\( U_k/f \)) balance will fail in any generalized theory of balance. As such, we now focus on a model valid near the equator that incorporates only this one difference.

\[2\text{ THE EQUATORIAL } \beta\text{-PLANE}\]

In standard coordinates the governing equations are

\[
\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} - \beta y v + \frac{\partial h}{\partial x} = 0, \tag{15}
\]

\[
\frac{\partial v}{\partial t} + u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + \beta y u + \frac{\partial h}{\partial y} = 0, \tag{16}
\]

\[
\frac{\partial h}{\partial t} + u \frac{\partial h}{\partial x} + v \frac{\partial h}{\partial y} + h \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) = 0, \tag{17}
\]

where \( (u, v) \) is the horizontal fluid velocity, \( \beta \) is the first derivative of the Coriolis parameter wr to \( y \), \( h \) is surface height and \( g \) is the constant acceleration due to gravity.

Linearizing about a rest state and assuming a non-linear frequency broadening the dispersion relation has solutions

\[
\omega_0 \sim Uk - \frac{\beta c^2 k}{(2n + 1)3c + 2Hk^2}, \tag{18}
\]

\[
\omega_\pm \sim Uk \pm \sqrt{(2n + 1)3c + 2Hk^2}, \tag{19}
\]

where \( U \) is some characteristic flow speed, \( H \) is the mean fluid depth, \( k \) is a wave number in the \( x \) direction and \( n \) is a non-negative integer indexing the Hermite polynomial factors of the solutions, which, in some sense, may be thought of as a squared wave number in the \( y \) direction. The ratio of the fast time scale to the slow time scale is

\[
\frac{\omega_0}{\omega_\pm} \sim \frac{1 + \frac{F}{R^2 + (2n + 1)F^2}}{1 + \frac{F}{R^2 + (2n + 1)F^2}} \tag{20}
\]
where \( R = U_\kappa / \sqrt{\beta c} \) is an equatorial Rossby number and
\( F = U / \sqrt{gH} \) is the Froude number. The situation is considerably
more complex now. Nevertheless, we can see immediately that larger values of \( n \) only make this quantity smaller. Assuming the worst case, we take \( n = 0 \). It follows that this ratio will be small if and only if the simultaneous conditions
\[
\epsilon_1 = \frac{RF}{\sqrt{R^2 + F^2}} \ll 1
\]
and
\[
\epsilon_2 = \frac{RF^2}{(R^2 + F^2)^{3/2}} \ll 1
\]
are satisfied.

Notice that
\[
\frac{\min[R, F]}{\sqrt{2}} \leq \epsilon_1 \leq \min[R, F]
\]
and
\[
\frac{1}{\sqrt{2}} \left( \frac{\min[R, F]}{\max[R, F]} \right)^2 \leq \epsilon_2 \leq \frac{\min[R, F]}{\max[R, F]}
\]
As in the \( f \)-plane case, small \( R \) or small \( F \) is equivalent to small \( \epsilon_1 \). However, this is not true for \( \epsilon_2 \). In fact, using
\( B = F/R \), we find
\[
\epsilon_2 = \frac{B^2}{(1 + B^2)^{3/2}}
\]
Evidently, we need one of \( F \) or \( R \) small but not both similarly small. For then, \( \epsilon_2 \) does not approach zero; we no longer have an asymptotic theory.

To understand this further, we consider the dispersion curves once again (figure 2). Now the slow motion includes waves with frequency \( \omega_R = -\beta c^2 k / (\beta c + c^2 k^2) \). It is \( \epsilon_1 \) small that minimizes the potential for frequency matching between the inertia-gravity waves and the vortical motion, exactly as in the \( f \)-plane case, while \( \epsilon_2 \) small does so between the inertia-gravity waves and Rossby waves. This is the motivation for our choice of the (equatorial) Rossby number: it is the ratio of the frequency of the vortical motion to the minimum frequency of the fast motion, \( \sqrt{\beta c} \). \( F \) is defined as before with similar interpretation. Note that Rossby waves are present at mid-latitudes but the ratio of the length scale of motion to the earth’s radius makes them much slower than inertia-gravity waves. At the equator \( f \) is zero and the Rossby waves must be included in the analysis.

Another striking difference is the appearance of the Kelvin wave solution. It is not clear whether this solution should be considered fast or slow. Up to this point we have been able to avoid specifying \( k \) in our interpretation of fast/slow motion, or, more precisely, assume moderate values of \( k \). The Kelvin wave dispersion curve, tangent to both that of the Rossby wave and the inertia-gravity wave, takes on both fast and slow frequencies at moderate values of \( k \).

Figure 2: The dispersion curves for the equatorial \( \beta \)-plane. \( \omega_G \) corresponds to \( \omega_k \) (inertia-gravity waves), \( \omega_K \) corresponds to Kelvin waves, \( \omega_V \) to \( \omega_0 \) (vortical motion) and \( \omega_R \) corresponds to Rossby waves (with a sign change for easy comparison).

Now, the analysis requires one single asymptotic parameter. We know \( B = F/R \) fails, essentially, because \( \epsilon_2(R, F) = \epsilon_2(B) \). An alternate function that suggests itself is
\[
m = \frac{\epsilon_2}{\epsilon_1} = \frac{F}{R^2 + F^2}
\]
It is justifiable to choose \( \epsilon_1 \) and \( m \) for a balanced theory. The resulting balanced equations are partially successful, failing only to capture the PG limit. It is not that this case is lost; it needs to be handled separately.

3. SUMMARY

A unified theory of balance on the sphere is well justified by our intuition and would seem quite possible. That is to say, the shallow water equations on a rotating sphere simply incorporate the physics of the above two simpler models. Indeed, the geometry is more complicated but we know the dispersion curves are similar. Moreover, it is only the PG limit that requires special treatment. Many circumstances, of practical importance restrict such a possibility. Nonetheless, the time scale separation technique seems to be the way ahead and holds the most promise to solving this problem.

5. REFERENCES