Mechanisms and Parameterizations of Geostrophic Adjustment and a Variational Approach to Balanced Flow

GEOFFREY K. VALLIS
University of California, Santa Cruz, California

(Manuscript received 18 November 1990, in final form 26 July 1991)

ABSTRACT

Geostrophic balance is shown to be the minimum energy state, for a given linear potential vorticity field, for small deviations of the height field around a resting state, in the shallow-water equations. This includes (but is not limited to) the linearized shallow-water equations. Quasigeostrophic motion is evolution on the slow manifold defined by advection of linear potential vorticity by the velocity field that minimizes that energy. Other linear and nonlinear arguments suggest that geostrophic adjustment is a process whereby the energy of a flow is minimized consistent with the maintenance of the potential vorticity field. A variational calculation that minimizes energy for a given potential vorticity field leads to a balance relationship that for the unapproximated shallow-water equations is similar but not identical to geostrophic balance. Preliminary numerical evidence, involving the inversion of potential vorticity for a simple model, indicates that this balance is a somewhat better approximation to the primitive equations than geostrophy.

It is also shown how the process of geostrophic adjustment may be significantly accelerated, or parameterized, in the primitive equations by the addition of certain terms to the equations of motion. Application of the parameterization to an unbalanced state in a primitive equation model is very effective in achieving a balanced state and in continuously filtering gravity waves. It is more accurate and less sensitive to tunable parameters than pure divergence damping, and may also be a useful and much simpler alternative to nonlinear normal-mode schemes whenever those may be inappropriate.

1. Introduction

Related questions of some importance in dynamical meteorology and oceanography are: Why are the ocean and atmosphere largely geostrophic? or Why are they "slow"? These questions motivate much of the work presented here, although no truly satisfactory answers are provided. The issues are intimately tied to the geostrophic adjustment process and the notion of balance—"slow" motion is evolution that is at least superficially devoid of gravity waves and approximately satisfies some balance relationship between velocity and pressure. Previous approaches to these issues have often relied on asymptotic theory and/or on a time-scale separation between gravity waves and geostrophic motion. In this paper we shall try to view the issue from a somewhat different perspective by arguing that geostrophic adjustment is a process of selective energy dissipation, and therefore that the subsequent slow motion should occur on or close to a minimum energy manifold. Indeed, we shall show that geostrophic balance is a minimum energy state for a certain set of primitive equations and then offer a new balance condition for the shallow-water equations based on a minimum energy principle rather than an asymptotic analysis. This picture of balanced flow leads naturally to a new method for rapidly achieving a balanced state, or parameterizing the geostrophic adjustment process, in the primitive equations.

Geostrophic balance (known at least since Buys Ballot 1857) is usually derived on the basis of a scale analysis. Taking the shallow-water equations as our primitive system, the $u$-momentum equation is

$$\frac{\partial u}{\partial t} + (u \cdot \nabla) u - f v = -g \frac{\partial h}{\partial x}, \tag{1.1}$$

with a similar equation for $v$. Notation is standard. Defining the Rossby number $R = U/f L$ where $L$ and $U$ are typical length and velocity scales, geostrophic balance appears to leading order in a Rossby number expansion. That is,

$$v \approx g \frac{\partial h}{f \partial x}, \tag{1.2}$$

and similarly for $u$. The quasigeostrophic system arises by advecting a linearized potential vorticity $\xi - f h/H$, where $\xi = k \cdot \nabla \times a$ is the vorticity, by the geostrophically balanced velocity field. Pedlosky (1979) and Veronis (1981) provide rigorous asymptotic derivations.

Corresponding author address: Dr. Geoffrey K. Vallis, Marine Science Department, University of California, Santa Cruz, CA 95064.

© 1992 American Meteorological Society
Use of such scaling arguments does not of course explain why the atmosphere should be geostrophic.

Even if the large scales are taken to be geostrophically balanced on average, it is not immediately apparent why there should not be continuous oscillatory gravity-wave motion around that state. A particular evolution of a primitive system will, for small Rossby number, settle down to a state apparently absent of gravity wave activity. The manifold on which subsequent motion occurs is often called the slow manifold (Leith 1980). It is now generally thought that for any nonzero Rossby number motion will not be completely devoid of gravity waves, although such activity may be of very small magnitude (e.g., Warn and Menard 1986). There appears to be no a priori reason why gravity wave activity should be very weak at large scales or, for that matter, why geostrophic balance should break down at small scales (Warn 1986). The theoretical calculations of Warn (1986) and the numerical results of Errico (1984) show that the slow manifold is actually unstable in the inviscid, spectrally truncated, limit. That is, motion that is geostrophically balanced at an initial time will, presuming ergodicity, subsequently develop strong gravity wave activity. The implication is that dissipation is necessary to keep the flow on the slow manifold.

Notwithstanding the nonexistence of a true slow manifold for the primitive equations, there has been much interest in deriving sets of equations that are devoid of gravity waves yet accurately represent the fluid motion of the atmosphere and ocean. Such sets are normally derived by asymptotic methods (assuming certain scales of motion) or by an adiabatic elimination based on the presence of “fast” variables (e.g., Lynch 1989; see van Kampen 1985 for a review of such methods). Approximations within a Hamiltonian framework have also been used (Salmon 1983), although this is also essentially a scaling approach within that framework. At low Rossby number it appears that the balance model provides a very good overall approximation, in terms of accuracy, to the primitive equations (Gent and McWilliams 1982; Allen et al. 1990; Barth et al. 1990).

Any process whereby some kind of balanced state is ultimately achieved [but not necessarily (1.2)] is usually called geostrophic adjustment. Rather than try to derive intermediate models based on an asymptotic analysis, a viable and physically based alternative procedure for deriving slow equations is to ask, What is the end state of the adjustment process?, without an a priori assuming geostrophic balance. In section 2 we show that geostrophic balance is in fact the minimum energy state for a given field of potential vorticity for a set of primitive equations that approximate the shallow-water equations. We argue in section 3 that geostrophic adjustment is in fact the process whereby the energy of an unbalanced state is reduced, either by gravity wave radiation to infinity or via an energy cascade to small scales where it is dissipated by viscosity, to the minimum consistent with the conservation of potential vorticity. In section 4, we use a minimization principle to derive a balance criterion similar to, but not exactly the same as, geostrophic balance. Section 5 presents some preliminary numerical evidence to support this conjecture. This picture of balanced flow then leads us, in section 6, to derive a rather simple parameterization for geostrophic adjustment that is able to damp gravity wave activity very effectively while barely affecting the slow motion and yet is much simpler than the machinery of nonlinear normal mode initialization. Section 7 concludes.

2. Geostrophic balance as a minimum energy state

In this section we show that geostrophic balance is a minimum energy state for approximations to the shallow-water equations with quadratic inviscid invariants. This includes the linearized shallow-water equations. Related (but different) results were obtained by Dikiy (1969) and Cullen et al. (1987) by quite different methods.

a. The shallow-water equations and their invariants

The shallow-water equations (e.g., Gill 1982) for a single layer of incompressible fluid with a free surface are

\[
\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} + f \mathbf{k} \times \mathbf{u} = -g \nabla h \tag{2.1}
\]

and

\[
\frac{\partial h}{\partial t} + \nabla \cdot (\mathbf{u} h) = 0. \tag{2.2}
\]

In these equations \( \mathbf{u} \) is the horizontal velocity \( \mathbf{u} = \mathbf{u} + j \mathbf{v} \), \( g \) is the acceleration due to gravity, and \( h \) is the height of the free surface. The first equation is a momentum equation, and the second arises from mass conservation.

It is convenient to write the velocity in terms of its divergent and rotational components; thus,

\[
\mathbf{u} = \nabla \phi + \mathbf{k} \times \nabla \psi, \tag{2.3}
\]

where \( \psi \) is a streamfunction and \( \phi \) a velocity potential. We also write \( h = H + h' \), where \( h' \) has zero mean. Making these substitutions the mass conservation equation becomes

\[
\frac{\partial h'}{\partial t} + J(\psi, h') + H \nabla^2 \phi + \nabla \cdot (\nabla \phi h') = 0. \tag{2.4}
\]

The momentum equations decompose into equations for the streamfunction and potential, namely:

\[
\frac{\partial}{\partial t} \nabla^2 \psi + J(\psi, \nabla^2 \psi) + \nabla \cdot (\nabla^2 \psi \nabla \phi) + f \nabla^2 \phi = 0 \tag{2.5}
\]
and
\[
\frac{\partial}{\partial t} \nabla^2 \phi + J(\phi, \nabla^2 \psi) + \frac{1}{2} \nabla^2 \left[ (\nabla \phi)^2 \right]
- \nabla^2 J(\phi, \psi) - \nabla \cdot (\nabla \psi \nabla \phi) + \frac{1}{2} \nabla^2 \left[ (\nabla \psi)^2 \right]
- f' \nabla^2 \psi + g' \nabla^2 h = 0. \tag{2.6}
\]

In the absence of forcing and dissipation the energy is conserved. That is,
\[
E = \int h(u^2 + v^2) + gh^2 \, dx, \tag{2.7a}
\]
\[
\frac{dE}{dt} = 0. \tag{2.7b}
\]

Further, it is easy to verify from (2.1) and (2.2) that potential vorticity is a material invariant in the sense that it is conserved on parcels. That is,
\[
\frac{Dq}{Dt} = 0, \tag{2.8}
\]
where \( Dq/Dt = \partial q/\partial t + (u \cdot \nabla) q \) and \( q = (\zeta + f)/h \).

Note that energy is a cubic invariant and that potential vorticity is a nonlinear Lagrangian invariant in the variables \( u \) (or \( \zeta \)) and \( h \).

The associated linear equations for the shallow-water system, obtained by a linearization about a state of rest and uniform height \( H \), are
\[
\frac{\partial u}{\partial t} + \beta x \times u = -g \nabla h', \tag{2.9}
\]
and
\[
\frac{\partial h'}{\partial t} + H \nabla \cdot u = 0. \tag{2.10}
\]

Written in rotational and divergent form, the momentum equations become
\[
\frac{\partial}{\partial t} \nabla^2 \psi + f \nabla^2 \phi = 0 \tag{2.11}
\]
and
\[
\frac{\partial}{\partial t} \nabla^2 \phi - f \nabla^2 \psi + g \nabla^2 h = 0. \tag{2.12}
\]

It is easily verified that the energy invariant of these equations is quadratic; that is,
\[
E = \int Hu^2 + gh'^2 \, dx, \tag{2.13}
\]
\[
\frac{dE}{dt} = 0. \tag{2.14}
\]

Potential vorticity conservation is now
\[
\frac{D}{Dt} \left( \zeta - f \frac{h}{H} \right) = 0. \tag{2.15}
\]

b. Geostrophic balance as a minimum energy manifold

We now show that geostrophic balance is obtained as the minimum energy for a given field of potential vorticity. This is a constrained problem in the calculus of variations, sometimes called an isoperimetric problem (Weinstock 1952), because of the origin of this class of problems in extremizing an area for a given perimeter. The energy to be minimized is given by (2.13) and the potential vorticity field by
\[
q = \zeta - f \frac{h}{H}. \tag{2.16}
\]

The constraint is incorporated by extremizing the integral
\[
I = \int H(u^2 + v^2) + gh'^2 + \lambda(x, y) \left( (v_x - u_y) - fh'/H \right) \, dx. \tag{2.17}
\]
The Lagrange multiplier is a function of space; if it were a constant, the integral would merely extremize energy subject to a given integral of potential vorticity, and rearrangements of potential vorticity would leave the integral unaltered.

There are three Euler–Lagrange equations obtained by minimizing \( I \). These are
\[
\frac{\partial \eta}{\partial h} - \frac{\partial}{\partial x} \frac{\partial \eta}{\partial h_x} - \frac{\partial}{\partial y} \frac{\partial \eta}{\partial h_y} = 0
\]
\[
\frac{\partial \eta}{\partial u} - \frac{\partial}{\partial x} \frac{\partial \eta}{\partial u_x} - \frac{\partial}{\partial y} \frac{\partial \eta}{\partial u_y} = 0
\]
\[
\frac{\partial \eta}{\partial v} - \frac{\partial}{\partial x} \frac{\partial \eta}{\partial v_x} - \frac{\partial}{\partial y} \frac{\partial \eta}{\partial v_y} = 0. \tag{2.18}
\]

where \( \eta \) is the integrand appearing in (2.17). Using (2.17) and (2.18) we obtain for the \( h \) equation
\[
2gh' - \lambda f = 0. \tag{2.19}
\]
The \( u \) and \( v \) equations yield
\[
2u + \frac{\partial \lambda}{\partial v} = 0
\]
\[
2v - \frac{\partial \lambda}{\partial x} = 0. \tag{2.20}
\]

Eliminating \( \lambda \) between (2.19) and (2.20) yields the simple relationships
\[
u = \frac{g \partial h'}{f \partial x}, \tag{2.21}
\]
which are otherwise known as the geostrophic balance relationships. Since (2.21) imply $\nabla \cdot u = 0$, we have

$$h' = f\psi/g.$$  \hspace{1cm} (2.22)

Therefore, we have proven that geostrophic balance is an extremum global energy state for a given field of linear potential vorticity. The energy extremum is clearly a minimum, since the rotational and divergent modes are orthogonal in the sense that

$$\int (k \times \nabla \psi + \nabla \phi)^2 = \int (\nabla \psi)^2 + (\nabla \phi)^2 \, dx,$$  \hspace{1cm} (2.23)

provided there are no boundary contributions (see also section 3b). Since $q = \nabla^2 \psi - fh'/H$, it is apparent that energy is always reduced by setting to zero the divergent modes associated with $\phi$ for a given $q$.

Quasigeostrophic evolution proceeds by allowing the potential vorticity to be advected by the geostrophic velocity. This is exactly the same as evolving the potential vorticity field (2.16) by the minimum energy velocity (2.21).

The global minimum of energy for a given integral of any function of potential vorticity for these equations is also a geostrophically balanced state. The problem is posed by extremizing the integral

$$I' = \int H \{u^2 + v^2\} + gh'^2$$

$$+ G(v - u - fh'/H) \, dx$$  \hspace{1cm} (2.24)

where $G$ is an arbitrary function of its argument. The resulting Euler–Lagrange equations are

$$2gh' - \frac{fG'}{H} = 0$$  \hspace{1cm} (2.25)

and

$$2Hu + \frac{\partial}{\partial y} G' = 0$$  \hspace{1cm} (2.26)

with a similar equation for $v$. From these we also obtain (2.21) as necessary conditions for energy extremization.

3. Mechanisms of geostrophic adjustment

In this section we argue physically and heuristically that the mechanism of geostrophic adjustment is one of energy dissipation and potential vorticity conservation. The arguments have their origins in the work of Rossby (1938), Sadourny (1975), and Warn (1986). Adjustment is the physical mechanism whereby a balanced state is achieved. If adjustment is a selective decay of energy, then since a minimum energy state is geostrophic, a dissipative system that nevertheless conserves potential vorticity will naturally evolve to a balanced state.

a. The linear adjustment problem

The canonical adjustment problem of Rossby, lucidly described by Gill (1982), makes simplifications of linearity, an infinite domain, and no viscosity. The initial state is one of rest with a step function in the height field, a manifestly unbalanced state. Because the problem is linear it may be solved analytically, although the physical aspect is clear. Gravity waves are excited, and a front propagates away from the discontinuity at a speed $\sqrt{gH}$, where $H$ is the fluid depth. Beyond the front the interface is undisturbed. Behind the front gravity wave activity takes place, and sufficiently far behind the front the fluid achieves a steady state. Now, the front serves to remove energy to infinity. However, the fluid cannot relax to a zero energy state because of the potential vorticity constraint, which here, because the problem is linear, is that potential vorticity conserved at each point. That is,

$$\frac{\partial}{\partial t} \left( \frac{H}{H} \right) = 0.$$  \hspace{1cm} (3.1)

Physically, then, the adjustment process proceeds by minimizing the energy constrained by local conservation of potential vorticity. The end state is a minimum-energy, geostrophically balanced state. The final configuration may thus be obtained by solving the simultaneous equations:

$$\nabla^2 \psi - \frac{fh}{H} = q_0$$

$$\psi - \frac{gh}{f} = 0$$  \hspace{1cm} (3.2)

where $q_0$ is the initial (and final) potential vorticity field.

b. Statistical equilibrium and nonlinear adjustment

Rossby's problem, being inviscid, requires an infinite domain to allow the divergent energy associated with gravity wave activity to escape. In a finite domain, unless viscosity is introduced, gravity waves will forever "slosh" without dissipating. In order to understand how viscosity might ultimately affect the adjustment problem, we briefly consider the inviscid statistical mechanical equilibrium.

In the limit of weak motion, the energy ($E$) and enstrophy ($Z$) invariants are quadratic. We nondimensionalize by measuring time in units of $f^{-1}$ and length in units of $(gH)^{1/2}/f$. Then, decomposing the fields into linear normal modes, the energy may be written:

$$E = \sum_k (E_G + E_V),$$  \hspace{1cm} (3.3)
where

\[ E_V = \frac{1}{2} \frac{|k^2 \Psi_k + h_k|^2}{1 + k^2} \]  

(3.4a)

is the potential-vortical energy and

\[ E_G = \frac{k^2}{2} \left( |\phi_k|^2 + \frac{|\psi_k - h_k|^2}{1 + k^2} \right) \]  

(3.4b)

is the divergent or gravity wave energy. The enstrophy has only a vortical-mode contribution and is

\[ Z = Z_V = \frac{1}{2} \sum_k |k^2 \psi_k + h_k|^2. \]  

(3.5)

(In geostrophic balance \( h_k = \psi_k \) and \( \phi_k = 0 \) and the divergent energy vanishes.)

The thermal equilibrium distributions are given by (Warn 1986)

\[ E_V = \frac{k}{(a + b \omega_k^2)} \]

\[ E_G = \frac{2k}{a} \]  

(3.6)

where \( \omega_k^2 = (1 + k^2) \) and we have again included the metric factor \( k \) in the numerator, as if the spectrum were continuous. These are sketched in Fig. 1. Energy in the gravity wave spectrum is equipartitioned among available modes, whereas the rotational or vortical modes feel the enstrophy constraint and have a distribution similar to that of incompressible two-dimensional flow (Kraichnan 1975). We can use these distributions to infer the sense of cascades in forced-dissipative problems. In the limit \( k_m \rightarrow \infty, a \rightarrow k_m^2, \) and \( b \rightarrow k_m^2 \), most of the energy is contained in the divergent part of the spectrum and close to the truncation wavenumber, although the vortical energy is still trapped at large scales. If the sense of nonlinear energy transfers is the same in forced-dissipative problems as it is in the approach to an inviscid equilibrium, then energy is transferred to gravity wave modes at high wavenumber, where it may be dissipated. The vortical modes, which feel the enstrophy constraint, remain trapped at larger scales and are presumably less affected by dissipation.

The numerical experiments of Farge and Sadourny (1989) by and large support the notion of energy dissipation during adjustment. Many of their integrations of the full shallow-water equations evidently have a direct cascade of divergent energy to small scales and a consequent selective dissipation of energy by viscosity.

\( c. \) Numerical evidence from a low-order model

We shall use a low-order model to see that during geostrophic adjustment gravity wave energy is indeed dissipated. The equations of motion are given by (2.4), (2.5), and (2.6). One of the simplest nontrivial implementations of these equations involves spectrally expanding the variables \( h, \phi, \) and \( \psi \) and truncating the resulting infinite set of ODEs to a single interacting triad. Since there are three variables and three components of the triad we obtain nine ODEs, which, following Lorenz (1980), are

\[ a_i \frac{dx_i}{dt} = a_i b_i x_j x_k - c(a_i - a_k) x_j y_k + c(a_i - a_j) y_j x_k - 2 c^2 y_j y_k + a_i (y_i - z_i) - \nu_0 a_i^2 x_i, \]  

(3.7)

\[ a_i \frac{dy_i}{dt} = -a_k b_i x_j y_k - a_j b_j y_j x_k + c(a_k - a_j) y_j y_k - a_i x_i - \nu_0 a_i^2 y_i, \]  

(3.8)
\[
\frac{dz_i}{dt} = -b_kx_j(z_k - h_k) - b_j(z_j - h_j)x_k + cy_j(z_k - h_k) - cy_k(z_j - h_j) + g_0a_i x_i - \kappa_0 a_i z_i + F_i. \tag{3.9}
\]

The variables \(x_i, y_j, z_i\) are the Fourier coefficients in the expansions of velocity potential \(\phi\), streamfunction \(\psi\), and free surface height \(h'\), respectively, and the indices \((i, j, k)\) take values \((1, 2, 3)\) cyclically, representing the three Fourier modes. The parameter \(a_i\) is the square of the \(i\)th wavenumber. The parameters \(b_i\) and \(c\) are functions of \(a_i\), and \(g_0\) is the square of the ratio of the deformation radius to the horizontal length scale. The parameters \(\kappa_0\) and \(v_0\) determine the level of dissipation. For a useful measure of the geostrophy of the problem, the Rossby number may be taken as the square root of \(F_1\) (McWilliams and Gent 1982). The parameter values we use are \(\kappa_0 = v_0 = 1/48, g_0 = 8, a_1 = a_2 = 1, a_3 = 3, h_1 = -1, h_2 = h_3 = F_2 = F_3 = 0\), and \(F_1\) determines the level of forcing, with typical values \(\sim 0.1\). (See Lorenz 1980; Gent and McWilliams 1982; and Warn and Menard 1985.)

A typical time evolution from unbalanced initial conditions is illustrated in Fig. 2. (Time stepping uses fourth-order Runge–Kutta, with a time step much less than that required for numerical stability.) Gravity waves are initially excited and slowly fade away, and the ultimate variability of the system is on a much longer time scale. Figure 3 shows that both energy and enstrophy fluctuate considerably during the adjustment process and in the subsequent slow evolution. To demonstrate the energy dissipation process we evaluate the linearized potential vorticity from the height and vorticity fields \(q = \zeta - fh' / H\) and then evaluate the geostrophic energy from this. The geostrophic energy is calculated by supposing that \(q\) consists only of a geostrophic contribution, so that \(q = \nabla^2 \psi - F \psi\) where \(E = f^2 / (gH)\) is related to the deformation radius. This equation is inverted to obtain \(\psi\), and the geostrophic energy is then calculated by

\[
E_g = \frac{1}{2} \int (\nabla \psi)^2 + F \psi^2 \, dx. \tag{3.10}
\]

The energy difference is the difference between this and (2.7a).

For our low-order model, the energy is given by (Gent and McWilliams 1982)

\[
E = 0.5 \sum (a_i x_i^2 + y_j^2) + z_k^2 / g_0 \]

\[-z_i(c y_j x_k - y_k x_j) / g_0 + b_i(x_j x_k + y_j y_k) \tag{3.11}\]

where the sum is taken over \((i, j, k) = (1, 2, 3)\) and cyclically. The terms on the last two lines form the nonquadratic contribution and are generally small (but not negligible) for most calculations reported here. The linear potential vorticity for each mode is proportional to \(q_i = a_i y_j + z_i / g_0\). We invert this to obtain the geostrophic streamfunction \(\psi_i = q_i / (a_i + 1 / g_0)\) and then obtain the geostrophic energy using

\[
E_g = 0.5[a_i(y_i^2) + (y_i^2) / g_0]. \tag{3.12}
\]

The difference between the geostrophic energy and true energy is plotted in Fig. 4 for two Rossby numbers, each beginning with unbalanced initial conditions of magnitude similar to those on the attractor. Initially, the energy difference is large, but it decays almost (but not exactly) monotonically. For small Rossby number the energy difference remains close to zero once the system has found the slow manifold, although for larger Rossby number the difference is somewhat larger. Note that the energy difference is not a positive definite quantity, although the difference between the energy

![Fig. 2. Time evolution of low-order primitive equation model from unbalanced initial conditions. The three curves are the \(x_1\) (velocity potential, long dashes), \(y_1\) (streamfunction, solid), and \(z_1\) (height, short dashes). In (a) (upper panel) \(F_1 = 0.1\). In (b) (lower panel) \(F_1 = 0.2\). In both cases gravity wave activity eventually fades away.](image)
calculated using only the terms on the first line of (3.11) and (3.12) is.

4. A generalized balance model

In the previous section it was argued that, from a physical standpoint, the adjustment process proceeds by reducing the energy constrained by the frozen field of potential vorticity. In section 2 it was shown that geostrophic balance is indeed the minimum energy state for a slightly reduced set of primitive equations. If geostrophic adjustment is an energy-minimizing process, then an appropriate balance will be the minimum energy state for the unapproximated shallow-water equations, constrained by the Erte potential vorticity field. Since most of the evidence presented in the previous section actually pertains directly only to weak flow this must for the moment remain a hypothesis. The variational problem is therefore to extremize the integral

$$ I = \int h(u^2 + v^2) + gh' + \lambda(x, y) \frac{\xi + f}{h} \ dx. \quad (4.1) $$

The resting potential energy $\int gh' / 2$ is unavailable for conversion to kinetic energy and hence dissipation. Hence, we should extremize only the available energy (i.e., kinetic plus available potential), and this is the reason for the prime on $h$ in the second term of the integrand. There is no assumption that $h' < H$.

Following the procedure of section 2b, the Euler-Lagrange equations (2.21) give rise to the simple balance relationships
\[ uh = - \frac{\partial}{\partial y} \left( \frac{B}{Q} \right) \]
\[ vh = \frac{\partial}{\partial x} \left( \frac{B}{Q} \right) \]  
(4.2)

where \( B \) is the Bernoulli function \( B = gh' + u^2/2 \) and \( Q \) is the potential vorticity \( Q = (\zeta + f)/h \). Since these conditions represent an extremal state, an appropriate name for them might be extreme balance. For small Rossby number the balance reduces to classical geostrophy. Details are given in appendix A. Note the shallow-water equations giving rise to (4.2) differ in two respects from the linearized equations: the potential vorticity is not linearized and the energy is not quadratic.

The extreme balance relationships (4.2) define only a balance; to obtain a model it must be decided how evolution along the slow manifold is to be achieved. One consistent choice is to advect Ertel potential vorticity by the balanced velocity. That is, a possible closed set of model equations is

\[ \frac{\partial Q}{\partial t} + (u \cdot \nabla)Q = 0, \]  
(4.3)

where the velocities are obtained from (4.2). Numerical integration of this model is made difficult by the implicit and nonlinear nature of the balance condition. From (4.2) we obtain

\[ \zeta = \nabla \cdot \frac{1}{h} \nabla \Psi \]  
(4.3)

where \( \zeta = Qh - f \) and \( \Psi = B/Q \). These nonlinear equations, with (4.2), must be solved iteratively and this may be difficult.

Note that extreme balance gives only the balanced velocity. In general we may suppose that the true velocity field may be decomposed into a balanced state and a secondary unbalanced flow; thus,

\[ v = v_b + v_s. \]  
(4.4)

The secondary flow may be obtained through the use of the balance condition and the equations of motion for the transport \( vh \) and Bernoulli function, analogous to the derivation of an omega equation in quasigeostrophic theory. This is done in appendix B.

5. Potential vorticity inversion by minimum energy

Given only the potential vorticity of the model described in section 3 it is possible to use the new balance conditions to obtain velocity and height fields that are at least better than a geostrophic inversion. Because of the nature of our numerical model, we actually invert linear potential vorticity; however, the results are encouraging. Our procedure is as follows. We integrate the model (3.7), (3.8), and (3.9). We form the potential vorticity from the height and vorticity values using \( q_i = a_i \nu_i + z_i/g_0 \). Given these values, we find the values of \( \{ x_i \}, \{ y_i \}, \) and \( \{ z_i \} \) that minimize energy given by (3.11) and preserve \( \{ q_i \} \). Rather than explicitly use the nonlinear balance equations (4.4), it is convenient to employ a numerical minimization algorithm; we employ a quasi-Newton method using a finite-difference gradient. For comparison purposes, and as a check on our algorithm, we also find the values of \( \{ x_i \}, \{ y_i \} \) and \( \{ z_i \} \) that minimize the quadratic form of the energy given by the first line of (3.11). This should simply give geostrophic balance, as indeed it turns out to do.

For small Rossby number, the minimum energy inversion performs quite well. In fact a time series of streamfunction values obtained for values of \( F_i \) less than about 0.25 is almost indistinguishable (by eye on a graph) from the true streamfunction. However, this is also true here for a geostrophic inversion. A more severe test is to look at the ageostrophy of the flow or the difference between the height and streamfunction fields. In Figs. 5 and 6 streamfunction is plotted in phase diagram against height. A geostrophic inversion

![Figure 4](image-url)  

**Fig. 4.** As in Fig. 2 except that energy differences (from geostrophic values) are plotted.
would yield a straight line through the origin, whereas the true phase diagram shows significant departures from this. These are quite well reproduced by the extreme balance, or energy minimization, scheme. This is encouraging evidence that the energy minimization principle is more than the merely linear argument that
the energy is comprised of orthogonal vortical and divergent modes and that only the vortical modes contribute to the potential vorticity. For higher values of Rossby number the primitive equations have quite strong gravity wave activity, and the inversion performs less well.
Two other aspects of the inversion should be mentioned. The first is that the value of the divergence is generally too small, by a factor of 2 or more. However, this is to be expected since only the balanced flow is determined by this inversion. A major contribution to the divergent flow will come from the secondary, unbalanced flow (appendix B). [If one were to perform an energy minimization on the linearized equations, or the quasi-primitive set of Farge and Sadourny (1989), then quasigeostrophy would result, known to be a good approximation for these equations. Yet divergence would be identically zero. Furthermore, divergence is not needed to integrate the quasigeostrophic potential vorticity equation.] The second aspect is an artifact of our model. If the velocity fields obtained from the inversion are to advect the model forward one time step, and then the potential vorticity is formed and inverted to obtain new velocity and height values and so on, the results are found to be little better than quasigeostrophy. This occurs because in the low-order model we invert a linearized potential vorticity; if the inverted fields were actually used to step the model forward in this way, we would thus be advecting the linear and not the true potential vorticity.

6. Practical use for gravity wave filtering

It is well known that if a numerical integration of the primitive equations is begun from more or less arbitrary initial conditions, gravity wave activity will initially be very high, damping over time to a much lower level as the system finds its slow (ish) manifold. Because the initial state may be unbalanced only because of observational or model inaccuracies, the fast motion is unacceptable from the point of view of numerical weather prediction. Cullen et al. (1987) also comment on the inadequacy of explicit geostrophic adjustment in primitive equation models during normal integrations and not only at the initial time, the implication being that some form of parameterization would be advantageous. A number of schemes have been devised to overcome these problems, among them nonlinear normal-mode initialization (Baer and Tribbia 1977; Machenauer 1977). Although very successful, if it had to be applied continuously, the procedure would be rather cumbersome. At the other extreme, divergence damping is an easily applied procedure to continuously damp gravity waves. However, it is not necessarily an accurate scheme if the damping is heavy.

The picture of slow motion drawn in the previous sections leads one naturally to suppose that a parameterization that would reduce the energy of a flow without directly affecting the potential vorticity should serve as a very simple but quite effective means of filtering unwanted gravity wave activity, or parameterizing geostrophic adjustment, in the primitive equations. In this section we relate this to divergence damping and show how to improve this by relaxing the divergence toward nonzero values determined by a balance criterion. But first we look at a conventional initialization using the balance model.

a. A balanced initialization

The first panel of Fig. 7 shows the evolution of (3.7)–(3.9) after the initial state has been balanced, that is, forced to satisfy the constraints of the balanced model. The balanced model consists of approximating the divergence equation (2.6) by

$$\nabla^2(\nabla \psi)^2/2 - \nabla \cdot (\nabla \psi \nabla^2 \psi) + g \nabla^2 h' - f \nabla^2 \psi = 0.$$  

For (3.7)–(3.9) the modified divergence equation (3.7) is

$$2 c^2 y_j y_k - a_i y_i + a_i z_i = 0, \tag{6.1}$$

thereby enforcing a balance relationship between streamfunction and height (Gent and McWilliams 1982). To achieve a balanced initial state we first enforce (6.1), choosing to do so in such a fashion that linear potential vorticity $q_i = a_i y_i + z_i / g_0$ is preserved. This is most easily done by an iterative process. For the initialization this procedure is implemented only at the first time step. By the presence of rather small oscillations visible to the eye in the divergence field, it can be seen that a balanced initialization is a good, but not perfect, method of eliminating gravity wave activity from a primitive equation integration.

b. Modifying the equations of motion

Vallis et al. (1989) showed how it is possible to modify the equations of motion to either generate or dissipate energy while still maintaining the topological invariants: in particular, potential vorticity and Kelvin's (or Bjerknes') circulation. Since geostrophic adjustment resembles such a process, we may expect that the adjustment process may be enhanced by such a parameterization.

An appropriate modification to the shallow-water momentum equation is

$$\frac{Du}{Dt} + f k \times u = -g \nabla h + \alpha \nabla h_i. \tag{6.2}$$

The last term on the right-hand side has the effect of dissipating or generating energy. That is,

$$E = \int \left[ gh^2 + hu^2 \right] dx, \quad \frac{dE}{dt} = -\int \alpha h_i \cdot \delta h_i dx.$$

However, there is no direct effect on the vorticity equation, and potential vorticity remains a material invariant. Note the consistency of this parameterization with the minimum energy criteria (4.2). From the mass conservation equation (2.2), $\partial h / \partial t = 0$ when $\nabla \cdot (uh) = 0$, which is implied by (4.2). Thus, with the sign of $\alpha$ appropriately chosen, Eqs. (6.1) and (2.2) form a system that will continuously dissipate energy until the minimum energy state $\nabla \cdot (uh) = 0$ is achieved.
Fig. 7. Evolution of the three components of height and streamfunction [left column, (a)-(c), streamfunction solid, height dashed] and divergence fields [right column, (d)-(f)] from unbalanced initial conditions with the parameterization (6.2). (a) and (d): $\alpha = 0$ but with balanced initialization. (b) and (e): $\alpha = 0.5$, no initialization. (c) and (f): $\alpha = 10$, no initialization. For no initialization and $\alpha = 0$, see Fig. 4.
Application of the modification to an initially unbalanced state reveals that it is quite effective in rapidly damping extraneous gravity wave activity, as illustrated in Fig. 7. The separation of time scales between gravity wave and geostrophic activity means that once the slow manifold has been reached the effects of the modification are slight. However, they are not completely negligible, as can be seen by increasing the value of $\alpha$. For higher and higher values, the evolution of the height field is constrained more and more until it becomes unrealistically small.

b. Relaxing to slow equations: Divergence damping and other schemes

Related parameterizations can be derived by relaxing the motion determined by the primitive equations toward some balanced manifold; that is, we accelerate convergence to the slow manifold by artificially relaxing the divergence to a (in general nonzero) value on that manifold. The simplest such parameterization is divergence damping in which the slow manifold is defined by the absence of divergence, or divergence tendency, that is, $\nabla \cdot u = 0$. Thus, the divergence equation becomes

$$\frac{\partial \delta}{\partial t} + \text{usual terms} = -\alpha \delta,$$

where $\alpha$ is a parameter and $\delta$ the divergence. This is a perhaps ad hoc but often useful method to operationally damp extraneous gravity waves in forecast models.

If we define the slow manifold by $\nabla \cdot h = 0$, then the slow manifold divergence, $\delta_s$, takes the value

$$\delta_s = -\frac{1}{h} (u \cdot \nabla h)$$

(6.3)

and the parameterization for the relaxation takes the form

$$\frac{\partial \delta}{\partial t} + \text{usual terms} = -\alpha \mathcal{L}(\delta - \delta_s)$$

$$= -\alpha \mathcal{L} \left[ \nabla \cdot u + \frac{1}{h} (u \cdot \nabla h) \right]$$

$$= -\alpha \mathcal{L} \frac{1}{h} \nabla \cdot (u h)$$

(6.4)

where $\mathcal{L}$ is some linear operator (e.g., a Laplacian) and $\alpha$ a constant whose sign is chosen appropriately. This is very similar to (6.2) and has similar properties of positive definiteness of energy decay.

A rather more accurate scheme than either of the above is to relax the divergence toward a slow manifold defined not by the absence of divergence tendency but by a balance criterion such as

$$\frac{\partial \delta}{\partial t} \{ g \nabla^2 h' - f \nabla^2 \psi \} = 0$$

(6.5)

or $\delta/\partial t = 0$, where $\epsilon$ is the "imbalance" $g \nabla^2 h' - f \nabla^2 \psi$. On the inertial manifold of the primitive equations the imbalance is expected to be very small, and the slow equations of Lynch (1989) are based around this hypothesis. The balance (6.5) is also imposed in the linear balance model, among other approximations.

From the shallow-water equations of motion we may derive an equation for the evolution of imbalance. This is

$$\frac{\partial \epsilon}{\partial t} + \text{nonlinear terms} = f^2 \delta - gH \nabla^2 \delta$$

(6.6)

or symbolically:

$$\frac{\partial \epsilon}{\partial t} + \text{n.l.} = \mathcal{L} \delta$$

(6.7)

where $\mathcal{L}$ is a linear operator and "n.l." represents the nonlinear terms. Defining a slow manifold by the absence of imbalance tendency, the appropriate divergence is given by

$$\mathcal{L} \delta_s = \text{n.l.}$$

(6.8)

We relax to the slow manifold by adding a term proportional to the difference between the true and slow divergences; thus,

$$\frac{\partial \delta}{\partial t} + \text{n.l.} = \alpha \mathcal{L} (\delta - \delta_s)$$

(6.9)

where $\mathcal{L}$ is the same as in (6.7). (If $\delta = 0$, the scheme is essentially equivalent to divergence damping.) Using (6.7) and (6.8) we find

$$\frac{\partial \delta}{\partial t} + \text{n.l.} = \alpha \frac{\partial \epsilon}{\partial t}$$

(6.10)

In a numerical model, the right-hand side of this would normally be evaluated diagnostically from the imbalance equation, with diabatic and viscous terms as needed. A number of variations on this theme are possible. Equation (6.10) has a couple of attractive features as a parameterization. First, it will reduce the energy of the flow when gravity waves are present. As Fig. 1 illustrates, the fast components of $\psi$ and $h$ are out of phase during the adjustment process, but their slow components track each other closely. Second, the parameterization is very insensitive to the value of $\alpha$. For $\alpha = 0$ the equations are just the usual, unmodified set. For $\alpha = \infty$ the equations are equivalent to the balance condition (6.5), which along with the unaltered vorticity and mass conservation equation forms an intermediate model with some of the features of both the linear balance system and the slow equations of Lynch. However, our purpose in this section is not to propose another ad hoc model but to accelerate convergence to a slow manifold. Figure 8 shows the evolution of (2.4), (2.5), and (6.10) from an unbalanced initial state for various values of $\alpha$. It can be seen to be re-
FIG. 8. As in Fig. 7 but using the scheme of (6.10). (a) and (d): $\alpha = 0.1$ and balance initialization. (b) and (e): $\alpha = 0.5$, no initialization. (c) and (f): $\alpha = 10$, no initialization.

Remarkably effective in removing gravity waves, while hardly disturbing the slow motion. This latter remark is justified by an integration beginning on the inertial manifold of the primitive equations; that is, the initial conditions are taken from values of the variables after a long integration of (3.7)–(3.9). For a very large range
of values of $\alpha$ the evolution of the parameterized model is indistinguishable from the original equations (Fig. 9). Indeed, the value of $\alpha$ in Fig. 9 is 100 times greater than that which is effective in eliminating gravity waves after a balanced initialization.

7. Discussion

In this paper we have argued that geostrophic adjustment is the process whereby the selective decay of energy occurs. The ultimate end state is one of minimum energy, constrained by the frozen potential vorticity field. Geostrophic balance was shown to be the minimum energy state for a system of primitive equations that has a quadratic energy invariant and a linear potential vorticity. We further showed that the minimum energy state for the true shallow-water equations, for a given potential vorticity field, is a generalized balance state that is normally close to, but not identical with, geostrophic balance. Some numerical evidence indicates that this may be a more accurate representation of the primitive equations than geostrophic balance. Finally, we presented a "parameterization" of geostrophic adjustment, and showed how effective it is in eliminating gravity waves and keeping a system close to a slow manifold. The practical advantage of such a scheme may lie in its simplicity of use, for example, making its implementation possible both as an initialization scheme or at every time step in situations where more complex schemes, such as nonlinear normal-mode initialization, might be inappropriate. For example, in nested primitive equation models it is easy to excite spurious gravity waves.

The schemes presented in this paper do not depend on the minimization of energy constrained only by the Lagrangian invariance of potential vorticity (i.e., the conservation of potential vorticity on parcels, allowing for the rearrangement of parcels). For if only that constraint were imposed, the extremal state of the system would be a steady, nonlinearly stable state. (Indeed, it is the essence of the energy–Casimir method of Arnold to find such states.) Rather, we propose that geostrophic adjustment occurs on a time scale shorter than that on which potential vorticity is advected, and therefore energy is minimized for a given ("frozen") field of potential vorticity.

On the theoretical side, the arguments for selective decay are rather heuristic; its occurrence may depend on the presence of two time scales, and therefore be essentially equivalent to more conventional approaches that exploit that fact in a more direct way. Of the intermediate models, extreme balance seems most closely related to the balance model. Both models conserve a true potential vorticity, and may be thought of as an advection of potential vorticity by a velocity field determined by certain balance criteria. Semigeostrophy (Hoskins 1975), on the other hand, advects a geostrophic field by a full velocity. Notwithstanding these differences and similarities, it should be noted that the goal of this paper is not so much to propose another intermediate model, but to try to approach the problem of balance from a physical rather than asymptotic perspective, seeking to understand why near geostrophy actually pertains in the atmosphere and ocean. If it is known a priori that the flow is at small Rossby number, say, then asymptotic approaches, constrained where possible by inviscid invariants, are naturally appropriate methods for model building (e.g., Allen 1991).

Acknowledgments. This work was supported by the NSF (ATM 8914004) and the ONR (N00014-90-J-1618). I am grateful to J. McWilliams for the asymptotic analysis of the balance state, T. Warn for many discussions, and the anonymous referees for their comments.

APPENDIX A

Asymptotic Analysis of Extreme Balance

Here we briefly analyze the extreme balance condition to ascertain its relationship with geostrophic balance at small Rossby number. This analysis was originally done by J. C. McWilliams.

Assuming a small Rossby number we asymptotically expand the height and velocity fields as

$$h = H + \epsilon h'$$  \hspace{1cm} (A.1)

$$u = \mathbf{k} \times \nabla \psi + \epsilon \nabla \phi,$$  \hspace{1cm} (A.2)

where $\psi$ is the streamfunction, $\phi$ a velocity potential,
and $\epsilon$ is the order Rossby number ordering parameter (and not the imbalance of $\S 6$).

The corresponding potential vorticity and Bernoulli fields are

$$q = \left( \frac{f}{H + \epsilon h} \right) \frac{f}{H} + \epsilon \left( \tilde{\xi} - \frac{fh}{H^2} \right) + O(\epsilon^2) \quad (A.3)$$

and

$$B = gh' + \frac{1}{2} \epsilon u^2. \quad (A.4)$$

From (4.2) we find to order $\epsilon$

$$\nabla^2 \psi + \epsilon \nabla \cdot \left( \frac{h'}{H} \nabla \psi \right) = \frac{g}{f} \nabla^2 h'$$

$$+ \epsilon \nabla^2 \left( \frac{1}{2f} (\nabla \psi)^2 + \frac{gh'}{2fH} - \frac{gh'}{f^2} \tilde{\xi} \right). \quad (A.5)$$

To lowest order it is clear that $h' = f \psi / g$, and substituting this into the order $\epsilon$ terms in (A.5) we find

$$f \nabla^2 \psi + \epsilon [2J(\psi_x, \psi_y) + \nabla \cdot (\psi \nabla[\tilde{\xi} - \psi/L^2])]$$

$$= g \nabla^2 h', \quad (A.6)$$

where $L^2 = gH/f$.

Therefore, to first order extreme balance is equivalent to geostrophy, and at the next order differs from gradient wind balance by the presence of the second term in square brackets in (A.6).

APPENDIX B

A Diagnostic for the Unbalanced Secondary Flow

Here we derive a diagnostic equation for the unbalanced component of the velocity field using extreme balance in the unforced, inviscid case. The equation is therefore analogous to the omega equation of quasi-geostrophic dynamics or an equation for the divergence in balanced systems. The procedure conventionally followed is to first write the equations of motion for vorticity, divergence, and height; one then uses a balance criterion in place of the full divergence equation; and, using this, time derivatives are eliminated between the vorticity and height equations. The procedure here differs only in detail.

Write the full transport (velocity times height) field as

$$V = vh = V_\phi + V_\rho \quad (B.1)$$

where $V_\phi$ is the balanced and rotational component of the transport field, given by

$$V_\phi = \hat{k} \times \nabla(\psi) \quad (B.2)$$

where $\psi = B/Q$, and $V_\rho$ is the residual, unbalanced flow.

Now, the evolution equations for the Bernoulli function (with $g = 1$) and the potential vorticity are

$$\frac{\partial B}{\partial t} + \nabla \cdot V + \frac{V}{h} \cdot \nabla B = 0 \quad (B.3)$$

and

$$\frac{\partial Q}{\partial t} + \frac{V}{h} \cdot \nabla Q = 0. \quad (B.4)$$

Therefore,

$$\frac{\partial \Psi}{\partial t} = -\frac{1}{Q} \left( \nabla \cdot V_\phi + \frac{V}{h} \cdot \nabla B \right) + \frac{B}{hQ^2} \nabla \cdot \nabla Q. \quad (B.5)$$

The evolution of $\nabla \times \Psi$ is given by

$$\frac{\partial}{\partial t} \nabla \times \Psi = \nabla \times \left[ \frac{V}{h} \nabla \cdot V + (V \cdot \nabla) \frac{V}{h} + f \hat{k} \times V \right]. \quad (B.6)$$

Equations (B.6), (B.5), and (B.2) are analogous to the more conventional vorticity, height, and balance equations, respectively. Eliminating time derivatives between (B.5) and (B.6) using (B.2) yields the omega equation:

$$\nabla \times \left\{ \frac{1}{h} (V_\rho + V_\phi) \cdot \nabla (V_\rho + V_\phi) \right\}$$

$$+ (V_\rho + V_\phi) \cdot \nabla (V_\rho + V_\phi)/h + f \hat{k} \times V_\phi \right\}$$

$$= -\nabla^2 \left[ \frac{1}{Q} \nabla \cdot V_\rho + \frac{1}{h} (V_\rho + V_\phi) \cdot \nabla B \right.$$

$$- \frac{B}{hQ^2} (V_\rho + V_\phi) \cdot \nabla Q \right\}. \quad (B.7)$$

Since $V_\rho$ is known in terms of $B$ and $Q$ this is an equation for the unknown $V_\phi$. It is highly implicit and nonlinear, and is likely to be difficult to solve. However, it is instructive to see that such an equation can in principle be derived. It will be simplified somewhat if the balanced flow alone is used in the advective terms in the evolution equations.

REFERENCES


Buys Ballot, 1857: Note sur la rapport de l'intensité du vent avec les


