Semigeostrophic Theory on the Hemisphere

GUDRUN MAGNUSDOTTIR AND WAYNE H. SCHUBERT

Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado

(Manuscript received 23 July 1990, in final form 18 December 1990)

ABSTRACT

This paper presents the combined isentropic and spherical geostrophic coordinate version of semigeostrophic theory. This is accomplished by first proposing a spherical coordinate generalization of the geostrophic momentum approximation and discussing its associated conservation principles for absolute angular momentum, total energy, potential vorticity and potential pseudodensity. We then show how the use of the spherical geostrophic coordinates allows the equations of the geostrophic momentum approximation to be written in a canonical form that makes ageostrophic advection implicit. This leads to a simple equation for the prediction of the potential pseudodensity. The potential pseudodensity can then be inverted to obtain the associated wind and mass fields. In this way the more general semigeostrophic theory retains the same simple mathematical structure as quasi-geostrophic theory—a single predictive equation which does not explicitly contain ageostrophic advection and an invertibility principle. The combined use of isentropic and spherical geostrophic coordinates is crucial to retaining this simplicity.

In order to demonstrate how the theory applies to problems of barotropic–baroclinic instability and Rossby–Haurwitz wave dispersion, we derive the semigeostrophic generalization of the Charney–Stern theorem and compare the semigeostrophic Rossby–Haurwitz wave frequencies with those of Laplace's tidal equations. The agreement between these frequencies is generally better than 0.5%. Thus, the theory appears to encompass a wide range of meteorological phenomena including both planetary-scale and synoptic-scale waves, along with their finer scale aspects such as fronts and jets.

1. Introduction

Semigeostrophic theory on the f-plane (Hoskins 1975; Hoskins and Draghici 1977), which combines the geostrophic momentum approximation and the transformation to geostrophic coordinates, has been widely and successfully used to study phenomena such as fronts, jets and the life cycle of baroclinic waves. Recent theoretical research has explored the possibility of extending semigeostrophic theory to include variable f. In particular, Salmon (1985) has used Hamilton's principle to generalize the definition of the coordinate transformation and the geostrophic balance to take account of a variable Coriolis parameter in a Cartesian framework. Shutts (1989) has extended this approach via Hamilton's principle to propose a planetary semigeostrophic theory. Using a more conventional approach that did not involve the use of Hamilton's principle, Magnusdottir and Schubert (1990) made use of Salmon's definitions to derive the geostrophic momentum approximation on the β-plane and the general form of β-plane semigeostrophic theory using the isentropic coordinate. In the present paper we again follow the more conventional approach and further generalize semigeostrophic theory to take account of the full variation of the Coriolis parameter on the sphere, defining a more general geostrophic balance and spherical geostrophic coordinates. Again, we use the isentropic coordinate in the vertical since combining it with the spherical geostrophic coordinates reduces the whole dynamics to just two equations—a predictive equation for the potential pseudodensity and a diagnostic equation (or invertibility principle) whose solution yields the balanced wind and mass fields from the potential pseudodensity. In isentropic and spherical geostrophic coordinates the divergent part of the circulation remains entirely implicit. The theory presented here has been developed as an extension of f-plane and β-plane semigeostrophic theory and is appropriate for the study of midlatitude phenomena when the full effects of the earth's sphericity are deemed necessary. It is subject to the same criticisms of its predecessors concerning the distortion of curvature effects and the neglect of the geostrophic advection of the ageostrophic flow. The theory is generally inappropriate for the study of equatorial phenomena and should thus be regarded as hemispheric rather than global. In this respect it should be considered a theory that is complementary to that of Shutts (1989), whose primary goal has been the development of a global extension of semigeostrophic theory.

Corresponding author address: Dr. W. Schubert, Department of Atmosphere Science, Colorado State University, Fort Collins, CO 80523

© 1991 American Meteorological Society

The outline of the paper is as follows. In section 2 we propose a generalization of the geostrophic momentum approximation to the sphere [Eqs. (2.1)–(2.7)] and show that this spherical geostrophic momentum approximation satisfies angular momentum and energy principles similar to the primitive equations, the primary difference being that the wind is evaluated geostrophically. Both the definitions of the geostrophic wind and the spherical geostrophic coordinates are generalizations of the β-plane definitions. We then derive the potential pseudodensity equation (section 3) and the invertibility relation (section 4), the combination of which constitutes the entire dynamics. In sections 5 and 6 we examine some consequences of linearized versions of the theory. We derive the generalized form of the Charney–Stern theorem on combined barotropic–baroclinic instability and compare the frequencies of semigeostrophic Rossby–Haurwitz waves with the frequencies obtained from Laplace's tidal equations.

2. Generalizations of the geostrophic momentum approximation and geostrophic coordinates

Using longitude λ, latitude ϕ, and potential temperature θ as independent variables, let us approximate the primitive equations by

$$\frac{Du_g}{Dt} - 2\Omega[v \sin\Phi + v_g(\sin\phi - \sin\Phi)] - \frac{u_g}{a}\left[v_g \tan\Phi + \left(v_g \frac{\cos\Phi}{\cos\phi} - v\right)\tan\varphi\right] + \frac{\partial M}{a\cos\phi\partial\lambda} = 0, \quad (2.1)$$

$$\frac{Dv_g}{Dt} + 2\Omega[u \sin\Phi + u_g(\sin\phi - \sin\Phi)] + \frac{u_g}{a}\left[u_g \tan\Phi + \left(u \frac{\cos\Phi}{\cos\phi} - u_g\right)\tan\varphi\right] + \frac{\partial M}{a\partial\phi} = 0, \quad (2.2)$$

$$\frac{\partial M}{\partial\theta} = \Pi, \quad (2.3)$$

$$\frac{D\sigma}{Dt} + \sigma\left(\frac{\partial u}{a\cos\phi\partial\lambda} + \frac{\partial(v\cos\phi)}{a\cos\phi\partial\phi} + \frac{\partial\dot{\theta}}{\partial\theta}\right) = 0, \quad (2.4)$$

where

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u\frac{\partial}{a\cos\phi\partial\lambda} + v\frac{\partial}{a\partial\phi} + \dot{\theta}\frac{\partial}{\partial\theta} \quad (2.5)$$

is the total derivative, $M = \theta\Pi + gz$ the Montgomery potential, $\Pi = c_p(p/p_0)^\kappa$ the Exner function, $\sigma = -\partial p/\partial\theta$ the pseudodensity, (u, v) the eastward and northward components of the velocity, (u_g, v_g) the geostrophic wind components given by

$$\left(v_g\frac{\cos\Phi}{\cos\phi}, -u_g\frac{\cos\phi}{\cos\Phi}\right) = \frac{1}{2\Omega\sin\Phi}\left(\frac{\partial M}{a\cos\phi\partial\lambda}, \frac{\partial M}{a\partial\phi}\right), \quad (2.6)$$

Φ the latitude part of the spherical geostrophic coordinates

$$(\Lambda, \sin\Phi) = \left(\lambda + \frac{v_g}{a2\Omega\sin\Phi\cos\Phi}, \sin\phi - \frac{u_g\cos\Phi}{a2\Omega\sin\Phi}\right), \quad (2.7)$$

and $\tan\varphi = (\sin\Phi + \sin\phi)/(\cos\Phi + \cos\phi)$. We shall henceforth refer to (2.1)–(2.7) as the spherical geostrophic momentum approximation. Equations (2.1) and (2.2) revert to the primitive equations when (u_g, v_g) are replaced by (u, v) and Φ is replaced by ϕ. Just as in the f-plane geostrophic momentum approximation the total time derivative acts only on the geostrophic part of the wind, which results in the same limitations regarding curvature effects as discussed in Hoskins (1975). The second parts of the Coriolis terms in (2.1) and (2.2) can be regarded as corrections for the fact that the first parts contain the Coriolis parameter evaluated at Φ rather than ϕ. A similar interpretation can be applied to the $\tan\varphi$ terms. The transformed latitude, the second part of (2.7), which enters the definition of the geostrophic wind (u_g, v_g) in (2.6), is an approximation to Hack et al.'s (1989) potential latitude coordinate. Because of the $\sin\Phi$ factor in the denominator of (2.6), there is difficulty in applying the theory globally. In the following we shall regard the theory as hemispheric and applicable primarily to midlatitude phenomena.

The motivation for the approximations (2.1)–(2.2) and the definitions (2.6)–(2.7) is that they collectively lead to the canonical momentum equations (2.18)–(2.19) and consequently to a form of the total derivative that does not have any horizontal ageostrophic advection. For typical atmospheric flows the positional shifts in transforming from (λ, ϕ) to (Λ, Φ) are small. For example, in a midlatitude region with geostrophic velocity components of order 10 m s^{-1}, the positional shifts are about 100 km.

The spherical geostrophic momentum approximation maintains important conservation principles of the primitive equations. The absolute angular momentum principle is obtained by simply noting that (2.1) can also be written as

$$\frac{D}{Dt}(\Omega a^2\cos^2\Phi) + \frac{\partial M}{\partial\lambda} = 0, \quad (2.8)$$

where, by (2.7), $\Omega a^2\cos^2\Phi = \Omega a^2\cos^2\phi + u_g a\cos\Phi \times (\sin\phi + \sin\Phi)/(2\sin\Phi)$ is the approximate absolute angular momentum. This shows that Φ is an approximate absolute angular momentum coordinate.

The kinetic energy principle associated with the spherical geostrophic momentum approximation can be obtained by adding u_g times (2.1) and v_g times (2.2). When this is done, there is cross-cancellation of the $\tan\Phi$ terms and the second part of the Coriolis terms.

Collecting the remaining terms and using (2.6) and (2.7) we obtain

$$\frac{DK_g}{Dt} + u\frac{\partial M}{a\cos\phi\partial\lambda} + v\frac{\partial M}{a\partial\phi} = 0,$$

where $K_g = \frac{1}{2}(u_g^2 + v_g^2)$ is the geostrophic kinetic energy. Combining this result with (2.4) we obtain

$$\frac{\partial}{\partial t}(\sigma K_g) + \frac{\partial}{a\cos\phi\partial\lambda}[\sigma u(K_g + gz)] + \frac{\partial}{a\cos\phi\partial\phi}[\sigma v\cos\phi(K_g + gz)] + \frac{\partial}{\partial\theta}\left[\sigma\dot{\theta}(K_g + gz) - gz\frac{\partial p}{\partial t}\right] + \sigma\alpha\omega = 0. \quad (2.9)$$

Multiplying (2.4) by c_pT we obtain the thermodynamic energy equation

$$\frac{\partial}{\partial t}(\sigma c_pT) + \frac{\partial}{a\cos\phi\partial\lambda}(\sigma u c_pT) + \frac{\partial}{a\cos\phi\partial\phi}(\sigma v\cos\phi c_pT) + \frac{\partial}{\partial\theta}(\sigma\dot{\theta}c_pT) - \sigma\alpha\omega = \sigma Q, \quad (2.10)$$

where $Q = \Pi\dot{\theta}$. The addition of (2.9) and (2.10) causes cancellation of the conversion term $\sigma\alpha\omega$ and leads to a total energy equation. We take the same view of the lower boundary as in Hoskins et al. (1985), assuming that isentropic surfaces cross the earth's surface, continuing just under it with pressure equal to the surface pressure. At any horizontal position where two distinct isentropic surfaces run just under the earth's surface, there is no mass trapped between them, so that $\sigma = 0$ there. Let us regard the bottom isentropic surface as the largest value of θ which remains everywhere below the earth's surface. Assuming the top boundary is both an isentropic and isobaric surface, assuming no topography and vanishing $\dot{\theta}$ at the top and bottom, we can integrate the total energy equation over the entire atmosphere to obtain

$$\frac{\partial}{\partial t}\iiint (K_g + c_pT)\sigma a^2\cos\phi d\lambda d\phi d\theta = \iiint Q\sigma a^2\cos\phi d\lambda d\phi d\theta. \quad (2.11)$$

Thus, except for the fact that the kinetic energy is evaluated geostrophically, the governing equations (2.1)–(2.7) have a total energy conservation principle identical to the one that exists for the primitive equations.

We will now proceed to transform (2.1) and (2.2) into the spherical geostrophic coordinates defined by (2.7). The purpose of the coordinate transformation is to make the ageostrophic flow completely implicit. Combining spherical geostrophic coordinates with the isentropic coordinate produces the desired result. Defining $\Theta = \theta$ and $T = t$ (but noting that $\partial/\partial\theta$ and $\partial/\partial t$ imply fixed λ, ϕ while $\partial/\partial\Theta$ and $\partial/\partial T$ imply fixed Λ, Φ), we can show that derivatives in $(\lambda, \phi, \theta, t)$ space are related to derivatives in $(\Lambda, \Phi, \Theta, T)$ space by

$$\frac{\partial}{\partial t} = \frac{\partial\Lambda}{\partial t}\frac{\partial}{\partial\Lambda} + \frac{\partial\Phi}{\partial t}\frac{\partial}{\partial\Phi} + \frac{\partial}{\partial T}, \quad (2.12)$$

$$\frac{\partial}{\partial\lambda} = \frac{\partial\Lambda}{\partial\lambda}\frac{\partial}{\partial\Lambda} + \frac{\partial\Phi}{\partial\lambda}\frac{\partial}{\partial\Phi}, \quad (2.13)$$

$$\frac{\partial}{\partial\phi} = \frac{\partial\Lambda}{\partial\phi}\frac{\partial}{\partial\Lambda} + \frac{\partial\Phi}{\partial\phi}\frac{\partial}{\partial\Phi}, \quad (2.14)$$

$$\frac{\partial}{\partial\theta} = \frac{\partial\Lambda}{\partial\theta}\frac{\partial}{\partial\Lambda} + \frac{\partial\Phi}{\partial\theta}\frac{\partial}{\partial\Phi} + \frac{\partial}{\partial\Theta}. \quad (2.15)$$

Applying (2.12)–(2.15) to the Bernoulli function $M^* = M + \frac{1}{2}(u_g^2 + v_g^2)$, it can be shown that

$$\left(\frac{\partial M}{\partial\lambda}, \frac{\partial M}{\partial\sin\phi}, \frac{\partial M}{\partial\theta}, \frac{\partial M}{\partial t}\right) = \left(\frac{\partial M^*}{\partial\Lambda}, \frac{\partial M^*}{\partial\sin\Phi} - \frac{u_g^2 + v_g^2(\cos^2\Phi - \sin^2\Phi)}{\sin\Phi\cos^2\Phi}, \frac{\partial M^*}{\partial\Theta}, \frac{\partial M^*}{\partial T}\right). \quad (2.16)$$

In most cases the additional term in the meridional transformation constitutes a small correction. An estimate of the ratio of the magnitude of this term to the magnitude of $\partial M/\partial\sin\phi$ is $(V/\Omega a)(1 + \cos 2\Phi)/(\sin\Phi \times \sin 2\Phi)$ where V denotes the typical magnitude of the geostrophic wind. For $V \sim 10$ m s^{-1} this ratio is 1/14 at 30°N, 1/75 at 60°N, and approaches zero at the pole. However, for strong flows such as the wintertime East Asian jet, this ratio may approach 1/2, in which case the additional term is not negligible. In section 4 we shall point out some simplifications which result from the neglect of this additional term.

The transformation relations (2.12)–(2.15) also imply that the total derivative (2.5) can be written as

$$\frac{D}{Dt} = \frac{\partial}{\partial T} + \frac{D\Lambda}{Dt}\frac{\partial}{\partial\Lambda} + \frac{D\Phi}{Dt}\frac{\partial}{\partial\Phi} + \dot{\theta}\frac{\partial}{\partial\Theta}. \quad (2.17)$$

With the aid of (2.16) we can now show that (2.1) and (2.2) take the canonical form

$$2\Omega\sin\Phi a\frac{D\Phi}{Dt} = \frac{\partial M^*}{a\cos\Phi\partial\Lambda}, \quad (2.18)$$

$$-2\Omega\sin\Phi a\cos\Phi\frac{D\Lambda}{Dt} = \frac{\partial M^*}{a\partial\Phi}. \quad (2.19)$$

When (2.18) and (2.19) are used in (2.17) for $D\Lambda/Dt$ and $D\Phi/Dt$, we see that a major advantage of the transformation from $(\lambda, \phi, \theta, t)$ space to $(\Lambda, \Phi, \Theta, T)$ space is the absence of ageostrophic advection in (2.17).

3. Fundamental prognostic equation: Potential pseudodensity equation

Here we shall use potential pseudodensity rather than potential vorticity as a fundamental variable because the theory in isentropic coordinates produces an elegant form of the potential pseudodensity equation. First we derive the equation for the absolute *isentropic* vorticity. Combining the derivatives of (2.18) and (2.19) in such a way as to form the total derivative of $2\Omega\sin\Phi\partial(\Lambda,\sin\Phi)/\partial(\lambda,\sin\phi)$, i.e., forming

$$\Lambda_\lambda[(2.18)\cos\Phi]_{\sin\phi} - (\sin\Phi)_{\sin\phi}[(2.19)/\cos\Phi]_\lambda - \Lambda_{\sin\phi}[(2.18)\cos\Phi]_\lambda + (\sin\Phi)_\lambda[(2.19)/\cos\Phi]_{\sin\phi},$$

results in

$$\frac{D\zeta}{Dt} + \zeta\left(\frac{\partial u}{a\cos\phi\partial\lambda} + \frac{\partial(v\cos\phi)}{a\cos\phi\partial\phi}\right) - \left(\xi\frac{\partial}{\partial\lambda} + \eta\frac{\partial}{\partial\sin\phi}\right)\dot\theta = 0, \quad (3.1)$$

where

$$(\xi, \eta, \zeta) = 2\Omega \times \sin\Phi\left(\frac{\partial(\Lambda, \sin\Phi)}{\partial(\sin\phi, \theta)}, \frac{\partial(\Lambda, \sin\Phi)}{\partial(\theta, \lambda)}, \frac{\partial(\Lambda, \sin\Phi)}{\partial(\lambda, \sin\phi)}\right). \quad (3.2)$$

Eliminating the horizontal divergence between (2.4) and (3.1) we obtain

$$\sigma\frac{DP}{Dt} = \left(\xi\frac{\partial}{\partial\lambda} + \eta\frac{\partial}{\partial\sin\phi} + \zeta\frac{\partial}{\partial\theta}\right)\dot\theta = \zeta\frac{\partial\dot\theta}{\partial\Theta}, \quad (3.3)$$

where $P = \zeta/\sigma$ is the Rossby–Ertel potential vorticity. The last step in (3.3) follows from (3.2) and (2.13)–(2.15). This shows that $\partial/\partial\Theta$ is actually the derivative along the vorticity vector and thus we could refer to $(\Lambda, \Phi, \Theta, T)$ as "vortex coordinates."

We define the potential pseudodensity σ^* by

$$\sigma^* = \frac{2\Omega\sin\Phi}{\zeta}\sigma, \quad (3.4)$$

so that the potential vorticity P and the potential pseudodensity σ^* are related by $P\sigma^* = 2\Omega\sin\Phi$. Thus, when (3.3) is manipulated into an equation for σ^* and when the total derivative is expressed in geostrophic space using (2.17)–(2.19), we can write the fundamental predictive equation of the model in the flux form

$$\frac{\partial\sigma^*}{\partial T} + \frac{\partial}{a\cos\Phi\partial\Lambda}\left(-\frac{\sigma^*}{2\Omega\sin\Phi}\frac{\partial M^*}{a\partial\Phi}\right) + \frac{\partial}{a\cos\Phi\partial\Phi}\left(\frac{\sigma^*}{2\Omega\sin\Phi}\frac{\partial M^*}{a\partial\Lambda}\right) + \frac{\partial}{\partial\Theta}(\sigma^*\dot\theta) = 0. \quad (3.5)$$

In the next section we shall discuss how the σ^* predictions of (3.5) can be inverted to obtain M^* and the associated wind and mass fields.

4. Fundamental diagnostic equation: Invertibility principle

The potential pseudodensity is a combination of the mass field σ and the wind field $2\Omega\sin\Phi/\zeta = \partial(\lambda, \sin\phi)/\partial(\Lambda, \sin\Phi)$. Since σ is related to M^* through hydrostatic balance and $\partial(\lambda, \sin\phi)/\partial(\Lambda, \sin\Phi)$ is related to M^* through geostrophic balance, σ^* depends only on M^*. Thus, all the balanced fields may be obtained from the potential pseudodensity through the invertibility principle. This relation between M^* and σ^* is derived as follows. From the definition of σ^* we have

$$\frac{\partial(\lambda, \sin\phi, \Pi)}{\partial(\Lambda, \sin\Phi, \Theta)} + \Gamma\sigma^* = 0, \quad (4.1)$$

where $\Gamma = d\Pi/dp = \kappa\Pi/p$. If the additional term in the second entry of (2.16) is neglected, the geostrophic relation for u_g simplifies, with the result that the geostrophic and hydrostatic relations in (Λ, Φ, Θ) space take the form

$$(fv_g, -fu_g, \Pi) = \left(\frac{\partial M^*}{a\cos\Phi\partial\Lambda}, \frac{\partial M^*}{a\partial\Phi}, \frac{\partial M^*}{\partial\Theta}\right), \quad (4.2)$$

where $f = 2\Omega\sin\Phi$. Using these geostrophic relations in the spherical geostrophic coordinate relations (2.7), we can express λ and $\sin\phi$ in terms of M^* as

$$\lambda = \Lambda - \frac{1}{f^2}\frac{\partial M^*}{a^2\cos^2\Phi\partial\Lambda}, \quad (4.3)$$

$$\sin\phi = \sin\Phi - \frac{\cos\Phi}{f^2}\frac{\partial M^*}{a^2\partial\Phi}. \quad (4.4)$$

Substituting (4.3), (4.4) and the last entry of (4.2) into (4.1), we obtain

$$\frac{1}{f^4}\begin{vmatrix} \dfrac{\partial^2 M^*}{a^2\cos^2\Phi\partial\Lambda^2} - f^2 & f^2\cos\Phi\dfrac{\partial}{a\partial\Phi}\left(\dfrac{1}{f^2}\dfrac{\partial M^*}{a\cos^2\Phi\partial\Lambda}\right) & \dfrac{\partial^2 M^*}{a\cos\Phi\partial\Lambda\partial\Theta} \\ \dfrac{\partial^2 M^*}{a^2\cos\Phi\partial\Lambda\partial\Phi} & f^2\dfrac{\partial}{a\cos\Phi\partial\Phi}\left(\dfrac{\cos\Phi}{f^2}\dfrac{\partial M^*}{a\partial\Phi}\right) - f^2 & \dfrac{\partial^2 M^*}{a\partial\Phi\partial\Theta} \\ \dfrac{\partial^2 M^*}{a\cos\Phi\partial\Lambda\partial\Theta} & \dfrac{\partial^2 M^*}{a\partial\Phi\partial\Theta} & \dfrac{\partial^2 M^*}{\partial\Theta^2} \end{vmatrix} + \Gamma\sigma^* = 0. \quad (4.5a)$$

Neglecting the effects of topography and assuming that the lower boundary is the constant height surface $z = 0$ and the isentropic surface $\Theta = \Theta_B$, we conclude that $M = \Theta\Pi$ at $\Theta = \Theta_B$. Written in terms of M^*, this lower boundary condition becomes

$$\Theta \frac{\partial M^*}{\partial \Theta} - M^* + \frac{1}{2f^2}\left[\left(\frac{\partial M^*}{a\cos\Phi\partial\Lambda}\right)^2 + \left(\frac{\partial M^*}{a\partial\Phi}\right)^2\right] = 0 \quad \text{at} \quad \Theta = \Theta_B. \tag{4.5b}$$

Assuming that the upper boundary is an isentropic and isobaric surface with potential temperature Θ_T and pressure p_T, the upper boundary condition for (4.5a) becomes

$$\frac{\partial M^*}{\partial \Theta} = \Pi(p_T) \quad \text{at} \quad \Theta = \Theta_T. \tag{4.5c}$$

The only lateral boundary condition required can be imposed at the equator or a conveniently chosen low latitude and depends on the particular application. For a given σ^*, we can regard (4.5) as a nonlinear second-order problem in M^*. Although Γ depends on M^*, this additional nonlinearity is weak.

Equations (3.5) and (4.5) form a closed system for the prediction of σ^* and the diagnosis of M^*. An efficient solver for (4.5) is a prerequisite since it must be solved at each time step. Fulton (1989) and Fulton and Schubert (1991) have described a multigrid solver for the two-dimensional f-plane problem. This solver addresses the problem of isentropes intersecting the earth's surface by adopting the massless region approach outlined in section 2. The discontinuity in σ^* on an isentropic surface also puts strict requirements on the numerical procedure for predicting σ^* using (3.5) since one might expect a ripple effect from the discontinuity. However, a problem similar to this one has been solved by Arakawa and Hsu (1990) and Hsu and Arakawa (1990). For solving (2.4) in a primitive equation model they proposed a finite difference scheme that has very small dissipation and computational dispersion, and guarantees positive definiteness. Thus, the massless layer approach to the lower boundary, which is consistent with the work of Bretherton (1966), seems to be useful not only conceptually but also computationally.

5. The Charney–Stern theorem generalized to semigeostrophic theory on the hemisphere

The Charney–Stern (1962) theorem is so fundamental that one would expect it to be valid for any consistent balanced theory. Here we derive the form of this theorem for semigeostrophic flow on the hemisphere. Our approach is similar to Eliassen's (1983) and to the approach used for β-plane semigeostrophic theory (Magnusdottir and Schubert 1990) and considers particle displacements around a mean flow. The mean flow is a vertically and meridionally varying zonal current. Neglecting frictional and diabatic effects, and linearizing the potential pseudodensity equation (3.5) around the mean flow, we obtain

$$\frac{\mathcal{D}\sigma^{*\prime}}{\mathcal{D}t} + v_g' f \frac{\partial}{a\partial\Phi}\left(\frac{\bar{\sigma}^*}{f}\right) = 0, \tag{5.1}$$

where $\mathcal{D}/\mathcal{D}t = \partial/\partial T - (1/f)(\partial\bar{M}^*/a\partial\Phi)\partial/a\cos\Phi\partial\Lambda$ and $fv_g' = \partial M^{*\prime}/a\cos\Phi\partial\Lambda$. Introduction of the northward geostrophic particle displacement η', defined by $v_g' = \mathcal{D}\eta'/\mathcal{D}t$, allows us to integrate (5.1) to obtain

$$\sigma^{*\prime} + \eta' f \frac{\partial}{a\partial\Phi}\left(\frac{\bar{\sigma}^*}{f}\right) = 0. \tag{5.2}$$

Multiplying (5.2) by v_g' and taking the zonal average at fixed Φ, we obtain

$$\frac{\partial}{\partial T}\left[\frac{1}{2}\overline{\eta'^2} f \frac{\partial}{a\partial\Phi}\left(\frac{\bar{\sigma}^*}{f}\right)\right] + \overline{v_g'\sigma^{*\prime}} = 0. \tag{5.3}$$

The linearized version of the invertibility relation (4.1) can be written as

$$\frac{\partial\Lambda'}{\partial\Lambda}\frac{\partial(\overline{\sin\phi}, \bar{p})}{\partial(\sin\Phi, \Theta)} + \frac{\partial((\sin\phi)', \bar{p})}{\partial(\sin\Phi, \Theta)} + \frac{\partial(\overline{\sin\phi}, p')}{\partial(\sin\Phi, \Theta)} + \sigma^{*\prime} = 0, \tag{5.4}$$

where $fa\cos\Phi\Lambda' = -v_g'$ and $fa(\sin\phi)' = u_g'\cos\Phi$. Multiplying (5.4) by v_g' and taking the zonal average, we obtain

$$\overline{v_g'\sigma^{*\prime}} = \overline{v_g'\left(\frac{\partial\bar{p}}{\partial\sin\Phi}\frac{\partial(\sin\phi)'}{\partial\Theta} - \frac{\partial\bar{p}}{\partial\Theta}\frac{\partial(\sin\phi)'}{\partial\sin\Phi}\right)} + \overline{v_g'\left(\frac{\partial\overline{\sin\phi}}{\partial\Theta}\frac{\partial p'}{\partial\sin\Phi} - \frac{\partial\overline{\sin\phi}}{\partial\sin\Phi}\frac{\partial p'}{\partial\Theta}\right)}.$$

Multiplying this by $f\cos\Phi$, rearranging with the aid of the zonally averaged thermal wind equation, and then substituting the result into (5.3), we obtain

$$\frac{\partial}{\partial T}\left[\frac{1}{2}\overline{\eta'^2} f^2\cos\Phi\frac{\partial}{a\partial\Phi}\left(\frac{\bar{\sigma}^*}{f}\right)\right] + \frac{\partial}{\cos\Phi\partial\Phi}\left[f\left(\frac{\partial\overline{\sin\phi}}{\partial\Theta}\overline{v_g'p'} - \frac{\partial\bar{p}}{\partial\Theta}\overline{v_g'(\sin\phi)'}\right)\cos\Phi\right] + \frac{\partial}{\partial\Theta}\left[f\left(\frac{\partial\bar{p}}{\partial\Phi}\overline{v_g'(\sin\phi)'} - \frac{\partial\overline{\sin\phi}}{\partial\Phi}\overline{v_g'p'}\right)\right] = 0, \tag{5.5}$$

which relates the time change of the wave activity to the divergence of the Eliassen–Palm flux. We now integrate (5.5) over the meridional plane, from the lower bounding latitude to the pole. The boundary flux at the top vanishes since both $\partial\bar{p}/\partial\Phi$ and p' vanish there. To show that the lower boundary flux vanishes, we proceed as follows. From the lower boundary condition on the basic state flow we have $\Theta_B\bar{\Pi} - \bar{M}^* + \bar{u}_g^2/2$

$= 0$. Differentiating this with respect to $\sin\Phi$ and using the geostrophic relation, we obtain

$$\Theta\bar{\Gamma}\frac{\partial\bar{p}}{\partial\Phi} + a2\Omega\tan\Phi\bar{u}_g\frac{\partial\overline{\sin\phi}}{\partial\Phi} = 0 \quad \text{at} \quad \Theta = \Theta_B. \quad (5.6)$$

From the lower boundary condition on the perturbation flow we have $\Theta_B\Pi' - M^{*\prime} + \bar{u}_g u_g' = 0$. Multiplying this by v_g' and taking the zonal average, we obtain

$$\Theta\bar{\Gamma}\overline{v_g'p'} + a2\Omega\tan\Phi\bar{u}_g\overline{v_g'(\sin\phi)'} = 0 \quad \text{at} \quad \Theta = \Theta_B. \quad (5.7)$$

Combining (5.6) and (5.7), we conclude that

$$\frac{\partial\bar{p}}{\partial\Phi}\overline{v_g'(\sin\phi)'} - \frac{\partial\overline{\sin\phi}}{\partial\Phi}\overline{v_g'p'} = 0 \quad \text{at} \quad \Theta = \Theta_B. \quad (5.8)$$

Thus, the lower boundary flux also vanishes and the result of integrating (5.5) over the meridional plane is

$$\frac{\partial}{\partial T}\iint\frac{1}{2}\overline{\eta'^2}f^2\cos^2\Phi\frac{\partial}{a\partial\Phi}\left(\frac{\bar{\sigma}^*}{f}\right)d\Phi d\Theta = 0, \quad (5.9)$$

and we conclude that the integral in (5.9) must be constant in time. In order for disturbances to grow in time, i.e., for $\overline{\eta'^2}$ to grow in time, $\partial(\bar{\sigma}^*/f)/\partial\Phi$ must have both signs. This is the spherical semigeostrophic generalization of the Charney–Stern theorem, which states that a necessary condition for combined barotropic–baroclinic instability is that the isentropic meridional derivative of the inverse potential vorticity must have both signs, i.e., potential vorticity must have an extremum along isentropes that slope from equator to pole.

6. Rossby–Haurwitz wave solutions

Consider adiabatic, frictionless flow for the quasi-Boussinesq case in which Γ is set equal to the constant $\Gamma_0 = R/p_B$. Then, linearizing about a basic state of rest with $\sigma_0 = (p_B - p_T)/(\Theta_T - \Theta_B)$, the linearized potential pseudodensity equation (5.1) and the linearized invertibility relation (5.4) can be combined into

$$\frac{\partial}{\partial T}\left\{\frac{\partial^2 M^{*\prime}}{\cos^2\Phi\partial\Lambda^2} + \frac{\sin^2\Phi}{\cos\Phi}\frac{\partial}{\partial\Phi}\left(\frac{\cos\Phi}{\sin^2\Phi}\frac{\partial M^{*\prime}}{\partial\Phi}\right) + \frac{4\Omega^2a^2\sin^2\Phi}{\Gamma_0\sigma_0}\frac{\partial^2 M^{*\prime}}{\partial\Theta^2}\right\} + 2\Omega\frac{\partial M^{*\prime}}{\partial\Lambda} = 0. \quad (6.1)$$

Searching for solutions of the form

$$M^{*\prime}(\Lambda, \Phi, \Theta, T) = \mathcal{M}(\Phi)\cos\left[\frac{\alpha_m(\Theta_T - \Theta)}{(\Theta_T - \Theta_B)}\right]e^{i(s\Lambda + \nu T)}, \quad (6.2)$$

we obtain the meridional structure equation

$$\mathcal{L}(\mathcal{M}) = \epsilon_m\mathcal{M}, \quad (6.3)$$

where

$$\mathcal{L} = \frac{d}{\cos\Phi d\Phi}\left(\frac{\cos\Phi}{\sin^2\Phi}\frac{d}{d\Phi}\right) + \frac{1}{\sin^2\Phi}\left(\frac{s}{\omega} - \frac{s^2}{\cos^2\Phi}\right), \quad (6.4)$$

$\omega = \nu/2\Omega$ is the nondimensional frequency, $\epsilon_m = 4\Omega^2a^2/c_m^2$ is Lamb's parameter, $c_m = c/\alpha_m$ and $c^2 = \Gamma_0\sigma_0(\Theta_T - \Theta_B)^2 = R(\Theta_T - \Theta_B)(p_B - p_T)/p_B$. The linearized lower boundary condition is satisfied if the constants α_m are the solutions of the transcendental equation $\alpha_m\tan\alpha_m = (\Theta_T - \Theta_B)/\Theta_B$. For the U.S. Standard Atmosphere we have $\Theta_T = 333$ K, $p_T = 22.5$ kPa, $\Theta_B = 287$ K, $p_B = 100$ kPa, in which case $c = 101.2$ m s^{-1}, and the solutions of the transcendental equation are $\alpha_m \approx 0.1241\pi$, 1.0160π, 2.0081π, 3.0054π, $\cdots$, corresponding to $m = 0, 1, 2, 3, \cdots$. The first of these roots ($m = 0$) corresponds to the external mode and yields $c_0 = 259.6$ m s^{-1} and $\epsilon_0 = 12.81$, while the remaining roots (which are approximately integral multiples of π) correspond to internal modes, the first of which yields $c_1 = 31.71$ m s^{-1} and $\epsilon_1 = 858.6$. For convenience of comparison of our results with the Laplace tidal equation results of Longuet-Higgins (1968), let us solve the eigenvalue problem (6.3) for $\epsilon_0 = 10$ and $\epsilon_1 = 1000$, which are characteristic of the external and first internal modes. We have discretized (6.3) on a uniform grid using second-order finite differences with $\Delta\Phi = 0.5$ deg and with boundary conditions $\mathcal{M} = 0$ at $\Phi = 0, \pi/2$. The resulting matrix eigenvalue problem has been solved for the eigenvalues ω and the corresponding eigenvectors using EISPACK routines. The eigenvalues for $s = 1, 2, \cdots, 5$ and for the first two meridional modes are shown in Table 1. For comparison, the values in parentheses give the percentage deviations of the semigeostrophic eigenvalues from Longuet-Higgins' (1968, Table 5) eigenvalues for those Rossby–Haurwitz modes with $\mathcal{M} = 0$ at $\Phi = 0$. In general, the semigeostrophic theory gives eigenvalues very close to those of Laplace's tidal equations, with a maximum error in Table 1 of 0.56%. This close agreement of semigeostrophic theory with tidal theory can be understood by noting that the tidal theory also leads to the eigenvalue problem (6.3) but with $\mathcal{L}$ defined by [Siebert 1961, Eq. (3.27); Matsuno 1966, Eq. (24); Longuet-Higgins 1968, Eq. (2.10); Chapman and Lindzen 1970, p. 110, Eq. (21)]

$$\mathcal{L} = \frac{d}{\cos\phi d\phi}\left(\frac{\cos\phi}{\sin^2\phi - \omega^2}\frac{d}{d\phi}\right) + \frac{1}{\sin^2\phi - \omega^2}\left(\frac{s(\sin^2\phi + \omega^2)}{\omega(\sin^2\phi - \omega^2)} - \frac{s^2}{\cos^2\phi}\right). \quad (6.5)$$

The only difference between the semigeostrophic operator (6.4) and the Laplace tidal operator (6.5) is the neglect of the ω^2 factors, which, according to Table 1, are quite small for Rossby–Haurwitz wave motion. This explains why the errors in the semigeostrophic eigenvalues become even smaller as ω becomes smaller.

TABLE 1. Semigeostrophic eigenvalues $\omega = \nu/2\Omega$ for $\epsilon_m = 10$ and 1000, first and second meridional modes, and zonal wavenumber $s = 1, 2, \cdots, 5$. The values in parenthesis are the percentage deviations of the semigeostrophic eigenvalues from the eigenvalues computed using the Laplace tidal equations (Longuet-Higgins 1968, Table 5). A negative percentage deviation means that the semigeostrophic eigenvalue is an underestimate of the Laplace tidal equation eigenvalue.

$\epsilon_m = 10$ (external mode)		$\epsilon_m = 1000$ (first internal mode)	
	$s = 1$		
.057793 (−0.40%)	.028366 (−0.04%)	.0063949 (−0.03%)	.0036298 (+0.04%)
	$s = 2$		
.082053 (−0.56%)	.042745 (−0.07%)	.012528 (−0.11%)	.0071675 (+0.02%)
	$s = 3$		
.088949 (−0.47%)	.049692 (−0.07%)	.018172 (−0.24%)	.010529 (−0.00%)
	$s = 4$		
.088402 (−0.34%)	.052626 (−0.06%)	.023161 (−0.37%)	.013643 (−0.04%)
	$s = 5$		
.084922 (−0.24%)	.053365 (−0.05%)	.027398 (−0.49%)	.016460 (−0.07%)

With such small errors in the Rossby–Haurwitz wave frequencies, one would expect the semigeostrophic model and the primitive equation model to give very similar simulations of Rossby–Haurwitz wave dispersion, with the obvious exception that the semigeostrophic model would not allow energy propagation across the equator.

In passing, it is interesting to note that the meridional structure equation (6.3) is very similar to one obtained by Matsuno (1970, 1971) in his heuristic extension of linear quasi-geostrophic theory to spherical coordinates for the study of wave propagation into the stratosphere.

7. Concluding remarks

The main results presented here are the generalized geostrophic momentum approximation and the spherical geostrophic coordinates (2.1)–(2.7), which together produce the two fundamental equations (3.5) and (4.5) for the prediction of potential pseudodensity σ^* and the diagnosis of M^* from σ^*. Thus, in isentropic and spherical geostrophic coordinates, the mathematical structure is the same as for semigeostrophic theory in isentropic coordinates on the f-plane (Schubert et al. 1989) and on the β-plane (Magnusdottir and Schubert 1990). In fact, we have encountered this same mathematical structure in previous studies of the gradient balanced vortex model of tropical cyclones (Schubert and Alworth 1987) and the zonal balanced model of the Hadley circulation (Schubert et al. 1990). The success of this general approach suggests that it may be the simplest way of looking at all types of balanced flows—whether they be planetary scale waves, synoptic scale midlatitude baroclinic waves with associated jets and fronts, or thermally forced tropical systems. The fact that the present theory is not global and does not encompass the tropical cyclone and Hadley circulation theories as special cases indicates that a more general theory should be sought. This more general theory would be global and would be based on a balance condition more general than geostrophy. This condition would include gradient balance as a special case, which would allow simulation of flows with large curvature vorticity. The formulation of the proper balance condition and coordinate transformation for such a theory remains a challenging and attractive problem. The attraction is that such a theory would be nearly all-encompassing. It would describe all balanced flows of meteorological interest with the exception of certain tropical systems which are not determined by potential vorticity dynamics. Work in this direction will be discussed in a forthcoming paper.

Acknowledgments. We would like to thank Glenn Shutts and Alan Thorpe for their helpful comments on generalizations of semigeostrophic theory. We would also like to thank Gerald Taylor for his mathematical advice and Paul Ciesielski and Piotr Flatau for their help with the solution of the eigenvalue problem. This work was supported by the National Science Foundation, under Grant ATM-8814541, and the Office of Naval Research, under Grant N000014-88-K-0214.

REFERENCES

Arakawa, A., and Y.-J. Hsu, 1990: Energy conserving and potential-enstrophy dissipating schemes for the shallow water equations. *Mon. Wea. Rev.*, **118**, 1960–1969.

Bretherton, F. P., 1966: Critical layer instability in baroclinic flows. *Quart. J. Roy. Meteor. Soc.*, **92**, 325–334.

Chapman, S., and R. S. Lindzen, 1970: *Atmospheric Tides.* Gordon and Breach, 200 pp.

Charney, J. G., and M. E. Stern, 1962: On the stability of internal baroclinic jets in a rotating atmosphere. *J. Atmos. Sci.*, **19**, 113–126.

Eliassen, A., 1983: The Charney-Stern theorem on barotropic-baroclinic instability. *Pure Appl. Geophys.*, **121**, 563–572.

Fulton, S. R., 1989: Multigrid solution of the semigeostrophic invertibility relation. *Mon. Wea. Rev.*, **117**, 2059–2066.

——, and W. H. Schubert, 1991: Surface frontogenesis in isentropic coordinates. *J. Atmos. Sci.*, **48**, in press.

Hack, J. J., W. H. Schubert, D. E. Stevens and H.-C. Kuo, 1989: Response of the Hadley circulation to convective forcing in the ITCZ. *J. Atmos. Sci.*, **46**, 2957–2973.

Hoskins, B. J., 1975: The geostrophic momentum approximation and the semi-geostrophic equations. *J. Atmos. Sci.*, **32**, 233–242.

——, and I. Draghici, 1977: The forcing of ageostrophic motion according to the semi-geostrophic equations and in an isentropic coordinate model. *J. Atmos. Sci.*, **34**, 1859–1867.

——, M. E. McIntyre and A. W. Robertson, 1985: On the use and

significance of isentropic potential vorticity maps. *Quart. J. Roy. Meteor. Soc.*, **111**, 877–946.

Hsu, Y.-J., and A. Arakawa, 1990: Numerical modeling of the atmosphere with an isentropic vertical coordinate. *Mon. Wea. Rev.*, **118**, 1933–1959.

Longuet-Higgins, M. S., 1968: The eigenfunctions of Laplace's tidal equations over a sphere. *Phil. Trans. Roy. Soc. London*, **A262**, 511–607.

Magnusdottir, G., and W. H. Schubert, 1990: The generalization of semigeostrophic theory to the β-plane. *J. Atmos. Sci.*, **47**, 1714–1720.

Matsuno, T., 1966: Quasi-geostrophic motions in the equatorial area. *J. Meteor. Soc. Japan*, **44**, 25–42.

——, 1970: Vertical propagation of stationary planetary waves in the winter northern hemisphere. *J. Atmos. Sci.*, **27**, 871–883.

——, 1971: A dynamical model of the stratospheric sudden warming. *J. Atmos. Sci.*, **28**, 1479–1494.

Salmon, R., 1985: New equations for nearly geostrophic flow. *J. Fluid Mech.*, **153**, 461–477.

Schubert, W. H., and B. T. Alworth, 1987: Evolution of potential vorticity in tropical cyclones. *Quart. J. Roy. Meteor. Soc.*, **113**, 147–162.

——, S. R. Fulton and R. F. A. Hertenstein, 1989: Balanced atmospheric response to squall lines. *J. Atmos. Sci.*, **46**, 2478–2483.

——, P. E. Ciesielski, D. E. Stevens and H.-C. Kuo, 1991: Isentropic potential vorticity modeling of the ITCZ and the Hadley circulation. *J. Atmos. Sci.*, **48**, in press.

Shutts, G. J., 1989: Planetary semi-geostrophic equations derived from Hamilton's principle. *J. Fluid Mech.*, **208**, 545–573.

Siebert, M., 1961: Atmospheric tides. *Advances in Geophysics*, Vol. 7, Academic Press, 105–187.