On "Too Fast" Baroclinic Planetary Waves in the General Circulation

W. K. DEWAR

Department of Oceanography, The Florida State University, Tallahassee, Florida

(Manuscript received 9 May 1997, in final form 4 December 1997)

ABSTRACT

Recent altimetric observations of the ocean surface reveal signatures of long planetary waves near the annual frequency band. Comparisons of the observed wave speeds and those predicted by standard linear theory suggest that the latter is inadequate as it yields different westward speeds than those measured; in the extratropical latitudes the predicted speeds are typically slower than the observations. Here the problem of long, baroclinic wave propagation in a forced, stratified ocean is considered theoretically with a view toward explaining these observations of "too fast" planetary waves.

From a quasigeostrophic analysis, it is argued that baroclinic waves in a sheared environment are accelerated to the west via their interactions with both the mean advective field and the mean potential vorticity field. Conditions under which the ratio of actual to linear phase speeds matches the observed ratio are computed and found to be typical of the open ocean. Extensions of these ideas to continuously stratified quasigeostrophic and layered planetary geostrophic systems are discussed.

1. Introduction

The wind and buoyantly driven general circulation is a highly time-dependent entity. In addition to internal modes of variability, for example, generated by mean flow instability, the large scale is subject to variable forcing in both momentum and heat fluxes. The ocean, when faced with changing conditions, responds with planetary waves that facilitate its adjustment. In many situations, these propagators take the form of long, nondispersive, baroclinic planetary waves of relatively low frequency (periods of 6 months and longer).

The recent advent of satellite-based altimetric measurements has for the first time yielded basinwide and global views of these all important waves. Early results in the analysis of the Ocean Topography Experimental Satellite Mission (TOPEX/Poseidon) data have been reported in Chelton and Schlax (1996, CS hereafter), who found that in the extratropical latitude bands (i.e., from $\pm 15°$ to $\pm 50°$) the observed propagation speeds of the sea level anomalies were roughly twice those of a classic linear planetary wave. Thus, CS judge standard linear theory wanting in terms of explaining their data. On the other hand, many of the premises of the linear theory seem more than apt. The time and space scales of the observations are consistent with geostrophy, and the surface signal displays the faster speeds at lower latitudes typical of the nondispersive planetary wave. Also, the neglect of relative vorticity seems eminently justified given the observed spatial scales (>500 km).

In this paper, we argue by means of theoretical analysis that augmenting the basic linear model in a relatively straightforward way with a mean, vertically sheared flow, in an ocean of realistic stratification, corrects the wave phase speeds from their linear value toward the observed values. The principal dynamical culprits behind this modification are wave interaction with the 1) background potential vorticity field, itself modified by mean flow, and 2) the vertical shear. We offer physical justifications for these effects and also extend the theory to the planetary geostrophic equations.

Background

The general observational problem of the behavior of long planetary waves, to which the above data are the most comprehensive to date, is one with a considerable precedent. Moving forward roughly in historical sequence, perhaps the first demonstration that North Pacific thermocline fluctuations could be interpreted as first-mode baroclinic Rossby waves is due to Bernstein and White (1974), who found good agreement between several east Pacific datasets and a simple wave model. Further evidence of such waves followed in Emery and Magaard (1976). White (1977) examined bathythermograph data from 11°N to 18°N and argued linear, nondispersive planetary wave dynamics in the presence

Corresponding author address: Dr. William K. Dewar, Department of Oceanography, The Florida State University, Tallahassee, FL 32306.
E-mail: bill@dva.ocean.fsu.edu

of deterministic forcing yielded an adequate model. Kang and Magaard (1980), using TRANSPAC XBT data, fit planetary wave dispersion relations computed both with and without mean shear. They concluded the variability was indeed representative of planetary waves and that the model with shear was marginally better at fitting the data.

Wind stress curl distribution was identified by White and Saur (1981) as a wave generating mechanism in the eastern Pacific, while Mysak (1983) argued for eastern current variability. Cummins et al. (1986), using an intermediate model, supported wind forcing as a dominant mechanism. Price and Magaard (1980) achieved mixed success modeling lower-frequency North Pacific variability as linear Rossby waves. Price and Magaard (1983) found roughly half of the 9–42-month open ocean variability observed along the Honolulu–San Francisco great circle could be accounted for in such a way. Closer to California, they found it necessary to account for shear. These same observations motivated Dewar (1989) to formulate the Doppler shift of the baroclinic waves by the time-dependent Sverdrup flow. Kessler (1990) examined an expanded North Pacific BT dataset, finding strong evidence for propagating Rossby waves and noting a systematic underestimation of the observed wave speeds by linear nondispersive planetary wave theory.

Price and Magaard (1986) performed a planetary wave fit to North Atlantic subsurface hydrographic data, but were able to explain only up to half (roughly) of the variability. They nonetheless concluded long planetary waves in the North Atlantic were indicated by their analysis. Krauss and Wuebber (1982) argued for time-dependent wind stress curls in the presence of the eastern boundary as a major wave generation mechanism in the North Atlantic. Lippert and Kase (1985) computed open ocean response to stochastic forcing, including a meridional boundary. This study was generalized by Hermann and Krauss (1989) to include realistic coastlines and bottom topography. Barnier (1988), in a two-layer quasigeostrophic numerical study, suggested the mid-Atlantic ridge divided the North Atlantic wave field into two separate domains. More recently, Sturges and Hong (1995) have concluded that low-frequency (periods greater than 3 yr) sea level variations at Bermuda are well predicted by a linear, nondispersive planetary wave model, in which local wave speed was adjusted according to the local stratification. Sturges et al. (1998) report similar findings for the U.S. East Coast at different latitudes.

Within this historical context, some of the CS results are classic. For example, they found wavelengths in the 0.5 to 2 cycles/year band ranging from 500 km at 50°N to 10 000 km at the equator (i.e., like those reported in several of the above studies) and noted a tendency for wave generation to occur near the eastern boundary. Chelton and Schlax also noted longitudinal dependency; phase speeds in the western half basins were routinely 50% greater than in the eastern half basin, and the waves were amplified in the west. The former result is new, while the latter is consistent with Bernstein and White (1977).

The question of wave speed is less clear from the historical analyses. Some studies argue for the adequacy of linear theory; others find the need to modify linear theory to allow for shear or slowly varying background stratification. Because of its unprecedented global coverage and high quality data, the CS analysis, however, lends considerable weight to the idea that pure linear theory is inadequate and that oceanic planetary waves move at speeds different from those of theoretical linear waves.[1] This strongly suggests that the linear theory of planetary wave propagation needs to be supplanted. Thus, Killworth et al. (1997) have revisited the effects of an existing mean shear on first-mode baroclinic propagation by considering a suitable extension of the eigenvalue problem. This study, which greatly supersedes those before it, argues that long-wave propagation in the presence of a baroclinic shear is generally augmented to the west by an amount roughly like the minimum value of the shear. This augmentation was computed from the Levitus dataset and found to generally correct the linear speed toward the observed values. Killworth et al. (1997) then include a barotropic correction and find an improved fit (although this is in part due in higher latitudes to an enlarged range of computed phase speeds). In summary, this study supports the conclusions of Kang and Magaard (1980) and Price and Magaard (1983) as to the relative importance of vertical shear on oceanic wave propagation.

In the present paper, we consider a setting like that analyzed by Killworth et al. (1997), although this approach is complementary. While the above authors have performed a local analysis (by means of solving a local eigenvalue problem), we here perform a global analysis. In addition, we extract the wave problem from the mean flow problem by capitalizing on the characteristic structure of the waves. Thus our approach, in principle, allows for all nonlinear interactions among all available modes, and the structure of the mean flow is computed as part of the solution rather than specified or taken from observations. The mechanisms leading to enhanced wave speeds are also presented in a somewhat more intuitive light. The costs of this analytical approach are that we work with idealized model stratifications (e.g., layers or a laterally homogenous basic state). In contrast, the eigenvalue problem solved by the above authors employed the observed stratification.

[1] We note that White (1977) predicted a factor of 2 increase of thermocline pattern speeds over the linear planetary wave propagation rate in his purely linear wave model. While this is surprisingly close to the amplification measured by CS, this result depends upon the variable wind forcing field obeying some structural constraints that are not characteristic of the data.

Also, the structure of our forcing (and thus mean) fields are relatively simple. Both of these costs can, in principle, be overcome if the calculation outlined herein is performed on a computer.

Accordingly, we find that the mean flow plays an important and surprising role in determining the propagation of first-mode waves and that nonlinear interaction among the baroclinic waves can be ruled out as a significant effect. Our model shows that the role of mean flow in affecting baroclinic waves is fourfold. Two of these roles are well known and consist of 1) Doppler shifting by the vertically averaged flow and 2) the modification of local wave production through alterations to the mean potential vorticity structure. The remaining two roles are less well understood and consist of 3) the interaction of the waves with the mean shear and 4) the feedback of the mean potential vorticity field on the propagation of waves. Interestingly, these tendencies operate in opposite senses: the Doppler shift (in item 1 above) operates in the sense of the circulation and the net mean shear effects (in items 3 and 4 above) operate in the sense opposite to the circulation. For reasonable parameters, we argue that the latter two effects dominate and can amplify wave phase speeds by order-one factors. It is interesting that this amplification occurs preferentially in the extratropical latitudes and may actually retard the waves in the lower latitudes. This is consistent with the published results of CS. Thus, we argue wave–mean flow interaction is a viable explanation for the CS observations of ''too fast'' baroclinic waves. In contrast, the present analysis is unable to explain the east–west amplitude differences noted by CS.

An analysis of a quasigeostrophic three-layer model is presented in the next section. Waves are generated in this model by means of a time-varying Ekman pumping, a mechanism which appears to be at least partly responsible for the waves observed by CS, and we analyze in particular the propagating waves. Analytical solutions in the limit of a deep third layer are compared with numerical solutions of the complete system in section 3. In section 4, extensions of the model to a continuously stratified system are derived. The physics of the model is discussed in section 5 and connections between the present theory and the many observations of planetary waves are made. An extension of the analysis to planetary geostrophic equations, which at large scales supersede quasigeostrophy, is given in the appendix (see also de-Szoeke and Chelton 1998).

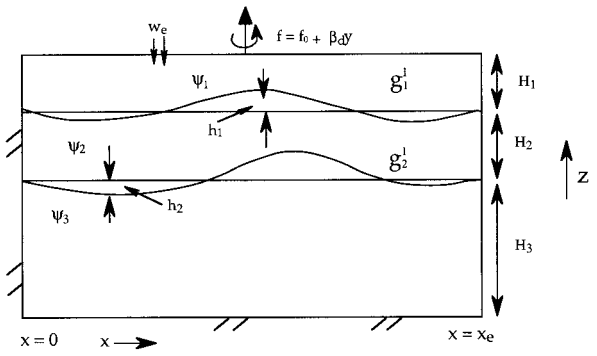

FIG. 1. Model schematic: a three-layer flat-bottom model forced at the surface by Ekman pumping. Notation is standard.

2. A three-layer quasigeostrophic model

We consider first the large-scale quasigeostrophic equations applied to a three-layer flat-bottomed system (see Fig. 1); the need for at least three layers will become apparent. The parametric restrictions on this system are that interface fluctuations are small and relative vorticity as a contribution to layer potential vorticity is neglected [see Pedlosky (1979) for a derivation]. Also, we employ a β plane, but the implicit latitudinal dependence of the Coriolis parameter is lost. Given the observed length scales of these waves, the neglect of relative vorticity does not pose a problem. More serious reservations arise when considering small interface fluctuations and a constant f_0. Relative to these, we here employ quasigeostrophy (QG) due to its simplicity. Extensions of these ideas to planetary geostrophic dynamics, which include a full Coriolis parameter and finite thermocline fluctuations, appear in the appendix. We are prepared to argue that much of the essential new dynamics of this problem are captured by the QG equations.

The three-layer QG equations are (see Fig. 1)

$$\frac{f_0^2}{g'H_1}(\psi_2 - \psi_1)_t + J\left[\psi_1, \beta_d y + \frac{f_0^2}{g'H_1}(\psi_2 - \psi_1)\right] = \frac{f_0 w_e}{H_1} \tag{2.1}$$

$$\frac{f_0^2}{g_1'H_2}(\psi_1 - \psi_2)_t + \frac{f_0^2}{g_2'H_2}(\psi_3 - \psi_2)_t + J\left[\psi_2, \beta_d y + \frac{f_0^2}{g_1'H_2}(\psi_1 - \psi_2) + \frac{f_0^2}{g_2'H_2}(\psi_3 - \psi_2)\right] = 0 \tag{2.2}$$

$$H_1\psi_1 + H_2\psi_2 + H_3\psi_3 = \frac{f_0}{\beta_d}\int_{x_e}^{x} w_e\, dx = \phi, \tag{2.3}$$

where the ψ_i are layer streamfunctions and other notation is standard. Equations (2.1) and (2.2) denote potential vorticity evolution in layers 1 and 2, and the system is closed by the Sverdrup constraint on net transport in (2.3).

We are explicitly interested in thermocline motion; thus, we eliminate layer streamfunctions in favor of the interfaces by the formulas $\psi_i - \psi_{i+1} = g_i' h_i / f_0$, where g_i' are the reduced gravities associated with interface i and a positive h_i denotes a depression in the thermocline. Use of the diagnostic in (2.3) then allows ψ_3 to be eliminated from the problem, yielding the two equations

$$-h_{1t} + J\left[\frac{\phi}{H} + \frac{g_1'}{f_0}\left(1 - \frac{H_1}{H}\right)h_1 + \frac{g_2'}{f_0}\left(1 - \frac{H_1 + H_2}{H}\right)h_2, \frac{\beta_d H_1}{f_0}y - h_1\right] = w_e \tag{2.4}$$

$$(h_1 - h_2)_t + J\left[\frac{\phi}{H} + \frac{g_2'}{f_0}\left(1 - \frac{H_1 + H_2}{H}\right)h_2 - \frac{g' H_1}{f_0 H}h_1, \beta_d \frac{H_2}{f_0} + h_1 - h_2\right] = 0, \tag{2.5}$$

where H is the total fluid depth. As we will later be interested in the limit of a very deep third layer (i.e., the classic reduced gravity limit), it is useful to nondimensionalize the above. Traditional QG scaling is used; that is,

$$h_i \sim \frac{f_0 UL}{g_i'}, \qquad \psi_i \sim UL \quad \text{for } i = 1, 2, \qquad x, y \sim L,$$

where L is the basin scale and U is a velocity scale. For convenience, Ekman pumping is scaled according to the Sverdrup relation; that is,

$$w_e \sim \frac{\beta_d U H_1}{f_0},$$

and the amplitude of ψ_3 is adjusted for the thickness of the third layer,

$$\psi_3 \sim \frac{H_1}{H_3}UL,$$

as suggested by (2.3). The result is

$$h_{1t} - J\left(\frac{H_1}{H}(\phi - h_1 - (1 + \lambda)h_2) + h_1 + h_2, \beta y - h_1\right) = -\beta w_e \tag{2.6}$$

$$(h_1 - \gamma h_2)_t + J\left(\frac{H_1}{H}(\phi - h_1 - (1 + \lambda)h_2) + h_2, \beta\lambda y + h_1 - \gamma h_2\right) = 0, \tag{2.7}$$

where all variables are now dimensionless and the primary system parameters are

$$\gamma = g_1'/g_2', \qquad \lambda = H_2/H_1 \quad \text{and} \quad \beta = \beta_d g_1' H_1/(f_0^2 U). \tag{2.8}$$

a. Linear analysis

It is useful to review the results of a purely linear analysis of (2.6) and (2.7). Neglecting all nonlinear terms yields (after some algebra)

$$h_{1t} - \beta\left(1 - \frac{H_1}{H}\right)h_{1x} + \beta\left(\frac{H_1}{H}(1 + \lambda) - 1\right)h_{2x} = -\beta\left(1 - \frac{H_1}{H}\right)\phi_x \tag{2.9}$$

$$-\gamma h_{2t} + \beta\left(1 - \frac{H_1}{H}(1 + \lambda)\right)h_{1x} + \beta(1 + \lambda)\left(1 - \frac{H_1}{H}(1 + \lambda)\right)h_{2x} = \beta\left(1 - \frac{(1 + \lambda)H_1}{H}\right)\phi_x. \tag{2.10}$$

The normal modes of the above are extracted by adding (2.9) and (2.10), the latter having been multiplied by an unknown coefficient, α. The aim of a normal mode analysis is to produce an equation in one variable only. Some straightforward algebra demonstrates that α values of

$$\alpha_{\pm} = \frac{-H}{2H_3\gamma}\left\{(1+\lambda)\frac{H_3}{H} - \gamma\left(\frac{H_2+H_3}{H}\right) \pm \left[\left((1+\lambda)\frac{H_3}{H} - \frac{\gamma(H_2+H_3)}{H}\right)^2 + 4\gamma\left(\frac{H_3}{H}\right)^2\right]^{1/2}\right\} \tag{2.11}$$

yield two equations of the form

$$(h_{\pm})_t - \beta_{\pm}(h_{\pm})_x = -\beta_{\pm}\phi_x, \tag{2.12}$$

where

$$\beta_{\pm} = \beta\left(\frac{H_2+H_3}{H} - \alpha_{\pm}\frac{H_3}{H}\right).$$

The two values of α denote the presence of two normal modes, representing the first and second baroclinic modes of this three-layer system. The "α_+" in (2.11) is identified with the first mode as it is seen to be negative, and thus $h_+ = h_1 - \alpha_+\gamma h_2$ measures the "in phase" movements of the interfaces. Conversely, $h_- = h_1 - \alpha_-\gamma h_2$ measures the "out of phase" movements.

We will adopt the form $w_e = a_1 a(y)\cos(\omega t)$ for the Ekman pumping, that is, a simple sinusoidal Ekman pumping at frequency ω and of amplitude a_1 whose spatial structure depends only on y. The lack here of a mean Ekman pumping is consistent with our linear analysis. The solution of (2.12) is

$$h_{\pm} = \frac{-2(\beta_{\pm})a_1 a(y)}{\omega}\sin\left(\frac{\omega(x_e - x)}{2\beta_{\pm}}\right) \times \cos\left[\left(\omega t - \frac{\omega(x_e - x)}{2\beta_{\pm}}\right)\right], \tag{2.13a}$$

where x_e denotes the location of the eastern boundary and consists of spatially modulated traveling waves. The factor of 2 appearing in the effective zonal wavenumber in (2.13) was first remarked upon by White (1977). The solution to (2.12) can also be written

$$\mathrm{h}_{\pm} = \frac{-(\beta_{\pm})a_1 a(y)}{\omega}\left\{\sin(\omega t) - \sin\left[\omega t - \frac{\omega}{\beta_{\pm}}(x_e - x)\right]\right\}, \tag{2.13b}$$

demonstrating that it consists of a locally forced response and a propagating part.

For all its simplicity, some important points emerge from the linear analysis. First, the length scales of the waves, $L_{\pm} = 2\pi\beta_{\pm}/\omega$, are different. Their ratio, for the values $\beta = \lambda = \gamma = 1$ and $H_1/H_3 = 0.1$, is

$$\frac{L_-}{L_+} = \frac{\beta_-}{\beta_+} \approx \frac{1}{6} < 1. \tag{2.14}$$

Second, both length scales are proportional to the timescale of the forcing, $T_s = 2\pi/\omega$. This relates to the first point in that the length scale of the waves generated against the eastern boundary is determined by the speed at which the waves propagate and the duration of a forcing cycle. These mechanisms are inherent in the formula for the wavelengths and underscore the result that the first- and second-mode waves travel at very different speeds. Last, it is seen in (2.13b) that the amplitudes of the modes scale as their wavelengths. Thus, the first mode is considerably stronger than the second mode with their relative amplitudes expressed by the ratio in (2.14). There is a good physical reason for this; namely, the forcing appears at the surface and is preferentially set up to generate a first-mode response.

An important comparison is to be made in this problem between the length scale of the basin and that of the generated waves. Writing the equation for the first-mode wavelength,

$$L_+ = \frac{2\pi\beta_+}{\omega} = \beta_d\left(\frac{H_2+H_3-\alpha_+H_3}{H}\right)\frac{g'H_1}{f_0^2U}T_s = \frac{T_sL}{U}\frac{\beta_d g'H_1}{f_0^2L}\left[\frac{H_2+H_3-\alpha_+H_3}{H}\right] = \frac{T_s^*}{T_b^*}, \tag{2.15}$$

where the starred quantities on the right are dimensional, reveals that the nondimensional wavelength can be related to the ratio of two timescales. The quantity $T_s^* = [2\pi/\omega]L/U$ is the timescale of the forcing and $L/\{\beta_d g'H_1 f_0^2[H_2+H_3-\alpha_+H_3]/H\}$ is seen to be the time necessary for a first-mode wave to cross the basin. Assuming an interest in the, say, annual frequency band (recall the CS waves were in the 0.5 to 2 cycles/year frequency band), L_+ is seen to be small because a typical midlatitude planetary wave transit time is roughly a decade; that is,

$$L_+ = \frac{1\text{ yr}}{10\text{ yr}} = 0.1 \ll 1. \tag{2.16}$$

Thus we arrive at the ultimate ordering in this precursor linear problem. The generated waves are found to have a small wavelength relative to the basin but with the first-mode waves longer than the second mode. Second, (2.13b) and (2.16) also demonstrate that while the first-mode waves are of large amplitude relative to the second mode, they are still of small amplitude relative to the adopted QG scaling.

b. Nonlinear analysis

It is the conclusion of CS that the linear first-mode wave phase speed β_+ given by (2.12) is too small by roughly a factor of 2 relative to observations for the latitude bands greater than (roughly) 20°. Thus we are motivated to move beyond the linear analysis of the previous section to the fully nonlinear problem. The

reader is warned, however, that we will capitalize on the fruits of the above linear analysis.

We also consider a more general forcing, w_e, which includes a mean Ekman pumping capable of driving a mean circulation. For definiteness, we will use

$$w_e = a(y)(a_0 + a_1 \cos\omega t), \tag{2.17}$$

although the analysis applies to considerably more general forcing fields. We emphasize two points here. First, no constraint to weak time-dependent forcing is assumed; that is, the variations in Ekman pumping can be comparable in strength to the mean field. This is useful as analyses of seasonal wind field changes argue they are a major perturbation to the overall wind structure forcing the oceans. Second, the fundamental constraint on both the steady and variable wind fields made here is that they are described by the basin scale. This is a comfortable assumption about the mean wind field, and also seems to adequately characterize the annual and lower-frequency variable wind field components [see Barnier (1986), who presents maps of the preferred North Atlantic wind field patterns associated with the annual frequency].

If a_1 is momentarily set to zero, the solutions of (2.6) and (2.7) are readily obtained using the methods of Rhines and Young (1982). Namely, flow in the lower layer vanishes where those layers are connected via characteristics to the eastern boundary, x_e. For inside areas isolated from x_e, potential vorticity is assumed to be uniform. Therefore, over large areas in the east, both lower layers are at rest (being exposed to x_e) and the steady solution of (2.6) and (2.7) is

$$h_{10} = \phi, \qquad h_{20} = 0, \tag{2.18}$$

where the subscript 0 denotes the mean field. Inside the isolated parts of layer 2, but exposed layer 3, regions,

$$\beta\lambda y + h_{10} - \gamma h_{20} = \beta\lambda y_N \tag{2.19a}$$

$$h_{10} + (1 + \lambda)h_{20} = \phi. \tag{2.19b}$$

The transition between (2.18) and (2.19) occurs if and when

$$\beta\lambda = -\phi_y \tag{2.20}$$

inside the basin. Finally, if conditions permit layer 3 to develop a closed pool of potential vorticity, the solutions for h_{10} and h_{20} are given by (2.19a), the conditions of uniform third-layer potential vorticity, that is,

$$\frac{\beta H_1}{H_3} y + \gamma h_{20} = \frac{\beta H_1}{H_3} y_N \tag{2.21}$$

and ψ_3 is given by

$$h_{10} + (1 + \lambda)h_{20} + \frac{H}{H_3}\psi_{30} = \phi. \tag{2.22}$$

The important result is that mean flow solutions are readily available, $O(1)$ in amplitude and possess basin-scale structure.

Now suppose both a_0 and a_1 are nonzero; that is, we consider the problem of the variably forced general circulation. We introduce the solution forms

$$h_1 = h_{10} + h_{11} \tag{2.23}$$

$$h_2 = h_{20} + h_{21}, \tag{2.24}$$

where h_{10} and h_{20} are defined by (2.18)–(2.22), into (2.6) and (2.7). The result, after some algebra, is

$$h_{11t} - \beta\left(\frac{H - H_1}{H}\right)h_{11x} - \frac{\beta H_3}{H}h_{21x} + J\left(\frac{H_1}{H}\phi, h_{11}\right) + J\left(\frac{H_1}{H}\phi', h_{10}\right) + J\left(\frac{H_3}{H}h_{20}, h_{11}\right) + J\left(\frac{H_3}{H}h_{21}, h_{10}\right) + J\left(\frac{H_3}{H}h_{21}, h_{11}\right) = -\beta\left(\frac{H - H_1}{H}\right)\phi'_x \tag{2.25a}$$

$$-\gamma h_{21t} + \frac{\beta H_3}{H}h_{11x} + \beta(1 + \lambda)\frac{H_3}{H}h_{21x} - \gamma J\left(\frac{H_1}{H}\phi, h_{21}\right) - \gamma J\left(\frac{H_1}{H}\phi', h_{20}\right) + \gamma\frac{H_1}{H}J(h_{10}, h_{21}) + \frac{\gamma H_1}{H}J(h_{11}, h_{20}) + \frac{\gamma H_1}{H}J(h_{11}, h_{21}) = \frac{\beta H_3}{H}\phi'_x, \tag{2.25b}$$

where $\phi' = \int_{x_e}^{x} aa_1 \cos(\omega t)\, dx$ is the variable part of the Ekman pumping.

The fully nonlinear equation for the first mode, $h_+ = h_{11} - \alpha_+\gamma h_{21}$, is now formed by combining (2.25a) with α_+ times (2.25b), where α_+ is defined by (2.11). We also use the fact that h_{11} and h_{21} may be expressed in terms of h_+ and h_-, namely,

$$h_{11} = a_{11}h_+ + a_{12}h_- \tag{2.26a}$$

$$h_{21} = a_{21}h_+ + a_{22}h_-, \tag{2.26b}$$

and the result is

$$h_{+t} - \beta_+ h_{+x} + J\left(\frac{H_1}{H}\phi, h_+\right) + J\left(\frac{H_1}{H}\phi', h_{10} - \alpha_+\gamma h_{20}\right) + J\left(h_{20}, \left(\frac{H_3}{H} - \alpha_+\gamma\frac{H_1}{H}\right)(a_{11}h_+ + a_{12}h_-)\right)$$
$$+ J\left(\left(\left(\frac{H_3}{H} - \alpha_+\gamma\frac{H_1}{H}\right)(a_{21}h_+ + a_{22}h_-), h_{10}\right) + (a_{21}a_{12} - a_{11}a_{22})\left(\frac{H_3}{H} - \alpha_+\gamma\frac{H_1}{H}\right)J(h_+, h_-)\right.$$
$$= -\beta\left[\frac{(H - H_1)}{H} - \frac{\alpha_+ H_3}{H}\right]\phi'_x. \tag{2.27}$$

Note that (2.27) contains terms that linearly and nonlinearly couple the first and second modes.

We now postulate that the scaling present in the linear problem persists into the nonlinear problem and seek solutions of the form

$$h_+ = L_+ h_+(x_+, y_+, t'), \qquad h_- = L_- h_-(x_-, y_-, t'), \tag{2.28}$$

where $L_\pm$ are the spatial scales of the linear waves appearing in (2.14) and $(x_\pm, y_\pm)$ are the ''short'' spatial coordinates associated with them. Further, the waves are both associated with the short temporal scale (ω^{-1}) due to their relatively high frequency forcing, and t' represents the associated temporal coordinate.

The above procedure explicitly recognizes the multiple space and time character of this problem. Mean flow is steady and large scale. The waves are in comparison high frequency and short. Within the wave response, the modes exhibit two space scales, which themselves are different. {While the factor 1/6 [see (2.14)] describing their scales is only marginally small, we are nonetheless motivated to examine a perturbation analysis.} Finally, the amplitude scaling for the (small) waves comes from the linear analysis. To complete the multiple scales approach to the problem, we rewrite the derivatives as

$$\frac{\partial}{\partial t} \to \frac{\partial}{\partial t} + \omega\frac{\partial}{\partial t'} \tag{2.29a}$$

$$\frac{\partial}{\partial x} \to \frac{\partial}{\partial x} + \frac{1}{L_+}\frac{\partial}{\partial x_+} + \frac{1}{L_-}\frac{\partial}{\partial x_-} \tag{2.29b}$$

$$\frac{\partial}{\partial y} \to \frac{\partial}{\partial y} + \frac{1}{L_+}\frac{\partial}{\partial y_+} + \frac{1}{L_-}\frac{\partial}{\partial y_-}. \tag{2.29c}$$

This is standard for a multiple-scales analysis and is described in Bender and Orszag (1978). Introducing the above in (2.27) and using (2.28) eventually yields

$$\omega L_+ h'_{+t} - \beta_+ h_{+x_+} + \frac{H_1}{H}\phi_x h_{+x} - \frac{H_1}{H}\phi_y h_{+y_+} + \frac{H_1}{H}[\phi'_x(h_{10} - \alpha_+\gamma h_{20})_y - \phi'_y(h_{10} - \alpha_+\gamma h_{20})_x]$$
$$+ \left(\frac{H_3}{H} - \alpha_+\gamma\frac{H_1}{H}\right)[h_{20x}(a_{11}h_{+y_+} + a_{12}h_{-y_-}) - h_{20y}(a_{11}h_{+x_+} + a_{12}h_{-x_-})]$$
$$+ \left(\frac{H_3}{H} - \alpha_+\gamma\frac{H_1}{H}\right)[(a_{21}h_{+x_+} + a_{22}h_{-x_-})h_{10y} - (a_{21}h_{+y_+} + a_{22}h_{-y_-})h_{10x}]$$
$$+ (a_{21}a_{12} - a_{11}a_{22})\left(\frac{H_3}{H} - \alpha_+\gamma\frac{H_1}{H}\right)(h_{+x_+}h_{-y_-} - h_{+y_+}h_{-x_-}) = -\beta\left(\frac{H - H_1}{H} - \frac{\alpha_+ H_3}{H}\right)\phi'_x, \tag{2.30}$$

where

$$\omega L_+ = 2\pi\beta\left(\frac{H_2 + H_3}{H} - \frac{\alpha_+ H_3}{H}\right) = 2\pi\beta_+ \sim O(1)$$

and ϕ' is the variable part of the Sverdrup transport.

As it stands, (2.30) represents a mix of a variety of spatial scales. These in turn measure nonlinear interactions occurring between the two baroclinic modes and the large-scale mean circulation. We now use the scale separation inherent in the formulation to simplify the above. This is achieved by averaging (2.30) over an area that is simultaneously large compared to the L_- scale and small compared to the L_+ scale. Note that this linear operation, when performed on (2.30), leaves all

but the derivatives dependent on x_- and y_- untouched, for the reason that h_{10}, h_{20}, and h_+ display no structure on the L_- scale and are thus sensibly constant over the area of integration. The result is

$$\omega L_+ h'_{+t} - \beta_+ h_{+x_+} + \frac{H_1}{H}[\phi_x h_{+y_+} - \phi_y h_{+x_+} + \phi'_x(h_{10} - \alpha_+\gamma h_{20})_y - \phi'_y(h_{10} - \alpha_+\gamma h_{20})_x]$$

$$+ \left(\frac{H_3}{H} - \alpha_+\gamma\frac{H_1}{H}\right)[h_{20x}(a_{11}h_{+y_+}) - h_{20y}(a_{11}h_{+x_+}) + a_{21}h_{+x_+}h_{10y} - a_{21}h_{+y_+}h_{10x}] + \beta\left(\frac{H_3 + H_2}{H} - \frac{\alpha_+ H_3}{H}\right)\phi'_x$$

$$= \frac{1}{A}\left\{\left(\frac{H_3}{H} - \alpha_+\gamma\frac{H_1}{H}\right)\left[h_{20y}a_{12}\int [h_-]\,dy_- - h_{20x}a_{12}\int [h_-]\,dx_- + a_{22}h_{10x}\int [h_-]\,dx_-\right.\right.$$

$$\left.\left. - a_{22}h_{10y}\int [h_-]\,dy_- + (a_{21}a_{12} - a_{11}a_{22})h_{+y_+}\int [h_-]\,dy_- - h_{+x_+}\int [h_-]\,dx_-\right]\right\}, \tag{2.31}$$

where the square brackets surrounding an integrand (e.g., $[h_-]$) denotes the evaluation of the integrand on opposite boundaries of the integration area. Thus the right-hand side of the equation consists only of line integrals, which scale as $A^{1/2}$, and is divided by the area of integration, A. Since the dimensions of the area are large compared to L_-, the right-hand side of (2.31) is $O(A^{-1/2}) \ll 1$, while the left-hand side is $O(1)$. Thus, the leading order behavior of the first baroclinic wave is obtained by neglecting the right-hand side of (2.31). This ''method of averaging'' approach is standard in the analysis of problems with many spatial scales and has been used in Pedlosky (1984) and Meng and Dewar (1995).

Returning to the original scaling of the problem, that is, that in use prior to the multiple-scales approach, (2.31) is written as

$$\underbrace{h_{+t} + J\left(\frac{H_1}{H}\phi + \beta_+ y,\, h_+\right)}_{\text{I}} + \underbrace{\left(\frac{H_3}{H} - \frac{\alpha_+\gamma H_1}{H}\right)[J(a_{11}h_{20} - a_{21}h_{10},\, h_+)]}_{\text{II}}$$

$$= \underbrace{-\beta\left(\frac{H_2 + H_3}{H} - \frac{\alpha_+ H_3}{H}\right)\phi'_x - J\left(\frac{H_1}{H}\phi',\, h_{10} - \alpha_+\gamma h_{20}\right)}_{\text{III}}. \tag{2.32a}$$

Thus first-mode wave propagation in the presence of a vertically sheared mean flow is governed by the normal westward drift due to β, Doppler shifted by vertically averaged mean flow (term I), interaction with the mean shear (term II) and local generation (term III). The latter involves the production by variable Ekman pumping of potential vorticity anomalies where the effect of the background mean field on the total potential vorticity is accounted for (the second part of term III).

Equation (2.32a) should be compared with (11) in Dewar (1989; see also Chang and Philander 1989), where the problem of wave propagation in a two layer flat-bottomed model was considered. It is thus seen that terms analogous to terms I and III occurred in that wave equation, which underscores term II as the principal new effect in this analysis. The fundamental difference between Dewar (1989) and this work is the present inclusion of a third layer, which, in turn, permits the vertical distribution of the mean flow to be different from that of the first baroclinic mode. Equivalently, in a two-layer model, the only degree of freedom available for constraining the vertical shear is the one available layer thickness. This being common to both the wave and the mean flow decouples the two, a result sometimes referred to as the non-Doppler effect. Held (1983) first pointed this out in a reduced

gravity model. Hermann and Kraus (1989), Chang and Philander (1989), and Kessler (1990) have added further discussion. Killworth et al. (1997) provide a somewhat more general demonstration, which will be returned to below. In the present problem, there are two degrees of freedom available to describe the vertical shear, and in general the structure of the mean shear will nontrivially affect the propagation of the first baroclinic wave mode. This interaction is what appears in term II in (2.32a); thus, we shall refer to term II as the wave–mean shear interaction term.

The above interpretation is reinforced upon evaluating a_{11} and a_{21} from (2.26). This allows (2.32a) to be written in the final form

$$\frac{\partial}{\partial t}h_+ + J\left(\frac{H_1}{H}\phi + \beta_+ y, h_+\right) + \left(\frac{H_3}{H} - \alpha_+\gamma\frac{H_1}{H}\right)\left(\frac{1}{\gamma(\alpha_+ - \alpha_-)}\right)J(h_{10} - \alpha_-\gamma h_{20}, h_+)$$
$$= -\beta\left(\frac{H_2 + H_3}{H} - \frac{\alpha_+ H_3}{H}\right)\phi'_x - J\left(\frac{H_1}{H}\phi', h_{10} - \alpha_+\gamma h_{20}\right), \quad (2.32b)$$

and it is noted that term II takes the form of first baroclinic mode wave interaction with the projection of the mean flow onto the second baroclinic mode, $h_{10} - \alpha_-\gamma h_{20}$.

The presence of the non-Doppler effect in the present three-layer problem is now clear. In this three-layer problem, there are three independent degrees of freedom represented by the barotropic mode and the first and second baroclinic modes. The interaction of the first baroclinic wave mode with the mean flow can therefore be expressed in complete generality by individual interactions of the wave with the reconstruction of the mean in terms of the three modes. In (2.32b), wave interaction with the mean barotropic mode appears in term I as a Doppler shift, and appears with the mean second baroclinic mode in term II; no interaction with the mean first baroclinic mode appears. The lack of such an interaction constitutes the non-Doppler effect in this problem. Killworth et al. (1997) find a similar result in their analysis of the eigenvalue problem.

A review of the analysis leading to (2.32) reveals wave–mean shear interaction consists of mean advection of anomalous wave potential vorticity (proportional to a_{11}) and wave advection of the mean field contribution to potential vorticity (proportional to a_{21}). Thus it is seen that wave propagation in the general circulation is influenced in two very different, although intuitively appealing, ways.

c. The reduced gravity limit

The effects of terms I and III were thoroughly explored in Dewar (1989). To focus clearly on wave mean shear interaction, we will for the remainder of the paper consider (2.32) in the limit of a very deep third layer. This is effected by allowing H_1/H and H_2/H to vanish, but $H_3/H \to 1$. The result is (dropping the subscript "+" on h_+)

$$h_t + J(a_{11}h_{20} - a_{21}h_{10} + \beta_+ y, h) = -\beta_+\phi'_x, \quad (2.33)$$

where now $\beta_+ = (1 - \alpha_+)$. It is seen that the Doppler shift and the mean potential vorticity feedback on wave generation are lost from the problem. Waves are thus influenced by simple local production and a net advective effect, with the latter consisting of a weighted sum of the second baroclinic mode of the mean field and the classic westward β drift.

d. Eastern basin propagation

In the east, $h_{20} = 0$ and $h_{10} = \overline{\phi}$ [where $\overline{\phi} = a_0\int_{x_e}^{x} a(y)\,dx$], so the advection field affecting the waves is simply

$$\chi = -a_{21}\overline{\phi} + \beta_+ y, \quad (2.34)$$

a plot of which appears in the eastern part of Fig. 2. In

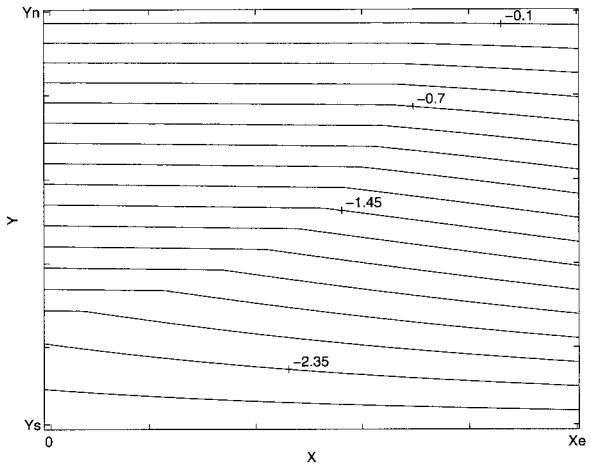

FIG. 2. First-mode wave characteristics in the deep third-layer limit. The eastern characteristics are given by (2.34) and the western by (2.43). The eastern and western regimes differ because of the western pool of uniform potential vorticity in the second layer.

the deep lower-layer limit, the quantity $a_{21} = -1/[\gamma(\alpha_+ - \alpha_-)]$. The quantities $\alpha_\pm$ have been evaluated from the form of (2.11) valid in the limit $H_3/H \to 1$ and using $\beta = \lambda = \gamma = 1$. The important, and surprising, point here is that the resultant mean field effect of the waves is entirely at odds with that which would be induced by a Doppler shift. In particular, the zonal wave speed induced by (2.34) is

$$-\chi_y = -\beta_+ + a_{21}\overline{\phi}_y = -\beta_+\left(1 - \frac{a_{21}\overline{\phi}_y}{\beta_+}\right). \quad (2.35)$$

Thus, the sense and magnitude of the correction depend upon a_{21}. From (2.11),

$$a_{21} = \frac{1}{\gamma(\alpha_- - \alpha_+)} = \frac{1}{((\lambda - \gamma + 1)^2 + 4\gamma)^{1/2}} > 0, \quad (2.36)$$

and thus the mean flow acts to retard westward waves in the south and accelerate them in the north. This cyclonic tendency is opposite to the otherwise anticyclonic circulation pattern of a subtropical gyre.

It is also straightforward to work out the ratio of the nonlinear to the linear wave phase speed; that is,

$$\frac{-\chi_y}{-\beta_+} = 1 - \frac{a_{21}\overline{\phi}_y}{\beta_+}. \quad (2.37)$$

The quantity $-\overline{\phi}_y$ is positive in the northern half of a subtropical gyre, so the effective wave speed is seen to be greater than the standard mean wave speed. It is instructive to express the order of the correction term using typical dimensional parameters. In doing this, recall that β_+ is the standard first baroclinic mode phase speed and, thus, when dimensionalized reads as $\beta_+ = \beta_d R_{d1}^2$, where R_{d1} is the first baroclinic deformation radius. The result is

$$r = \frac{-a_{21}\overline{\phi}_y}{\beta_+} = 0\left(\frac{w_e f_0}{\beta_d(\beta_d R_{d1}^2)H_1((\lambda - \gamma + 1)^2 + 4\gamma)^{1/2}}\right). \quad (2.38)$$

Typical Sverdrup scaling suggests $w_e \sim \beta_d U H_1/f_0$, and it is also true that the standard long-wave phase speed is comparable to U. Finally, $[(\lambda - \gamma + 1)^2 + 4\gamma]^{1/2}$ is made up of $O(1)$ numbers and is thus also $O(1)$. Thus, (2.38) shows the potential westward acceleration of baroclinic waves by their interaction with the mean shear is the same order as the standard linear phase speed. If the reasonable parameters $H_1 = 250$ m, $H_2 = 500$ m, $g' = 0.014$ m s^{-2}, $g'' = 0.007$ m s^{-2}, $\beta_d = 2 \times 10^{-11}$ m^{-1} s^{-1}, and $f_0 = 10^{-4}$ s^{-1} are chosen, the ratio r in (2.38) is unity for $w_e = 2 \times 10^{-6}$ m s^{-1}. This value corresponds to maximum Ekman pumping values found in the North Atlantic (Leetmaa and Bunker 1978) and Pacific Oceans (Gill 1982).

Of course, the above estimate does not account for geographic variation of the effect of the mean flow. For example, evaluating $\overline{\phi}_y$ yields

$$r = \frac{a_{21}}{\beta_+}a_0\left(\frac{\partial}{\partial y}a\right)(x - x_e), \quad (2.39)$$

demonstrating that the wave acceleration vanishes on the eastern boundary and grows in strength with increasing westward displacement from x_e.

This calculation shows that ratios of 2 for the observed to predicted linear wave speeds for the northern part of the general circulation are to be expected. It also suggests ratios of less than unity should be found to the south. The former tendency is clearly indicated in Fig. 5 of Chelton and Schlax (1996). The latter, having received less attention, is nonetheless present in those same results.

Analytical solutions for the waves are available for the choice

$$w_e = \begin{cases} (y - 1)(1 + \cos(\omega t)), & y > 1/2 \quad (2.40a) \\ -y(1 + \cos(\omega t)), & y < 1/2, \quad (2.40b) \end{cases}$$

that is, time-dependent Ekman pumping varying linearly with latitude. The form of (2.33) is that of a first-order equation for which the method of characteristics is appropriate. The characteristics equations are

$$t_s = 1 \quad (2.41a)$$

$$x_s = \frac{-(x - x_e)}{\gamma(\alpha_+ - \alpha_-)} - \beta_+ \quad (2.41b)$$

$$y_s = \begin{cases} \dfrac{(y - 1)}{\gamma(\alpha_+ - \alpha_-)}, & y > 1/2 \quad (2.41c) \\ \dfrac{-y_-}{\gamma(\alpha_+ - \alpha_-)}, & y < 1/2 \quad (2.41d) \end{cases}$$

$$h_s = \begin{cases} -\beta_+(y - 1)\cos(\omega t), & y > 1/2 \quad (2.41e) \\ \beta_+ y \cos(\omega t), & y < 1/2 \quad (2.41f) \end{cases}$$

to which the boundary conditions $t = r$, $h = 0$, $y = y_0$, $x = x_e$ are applied. The solutions for x and h in the northern half plane are

$$x = x_e - \beta_+\gamma(\alpha_+ - \alpha_-)(1 - e^{-s/[\gamma(\alpha_+ - \alpha_-)]}) \quad (2.42a)$$

$$h = \frac{\beta_+(y - 1)\gamma(\alpha_+ - \alpha_-)}{[1 + \omega^2\gamma^2(\alpha_+ - \alpha_-)^2]} \times (\cos(\omega t) + \omega\gamma(\alpha_+ - \alpha_-)\sin(\omega t)) - \frac{\beta_+(y_0 - 1)\gamma(\alpha_+ - \alpha_-)}{[1 + \omega^2\gamma^2(\alpha_+ - \alpha_-)^2]} \times (\cos(\omega t_0) + \omega\gamma(\alpha_+ - \alpha_-)\sin(\omega t_0)), \quad (2.42b)$$

where t_0 is the wave ''emission'' time from the eastern boundary and $s = t - t_0$ is time since emission. Thus, (2.42b) shows h consists of a local response proportional

to absolute time t and a propagating wave proportional to emission time t_0. Of principal interest to this study is the propagating wave, and (2.42a) demonstrates the increasing role of the mean flow effect with increasing distance from x_e. A similar set of results is found for the southern half basin, that is, $y < 1/2$.

e. Western basin propagation

Farther west in the basin, the second layer in this model is put into motion by Rhines and Young mechanics. McDowell et al. (1983), Keffer (1985), and Talley (1988) provide observational support for the existence of the related uniform potential vorticity pools. In such an area, the net mean field effect is augmented by a contribution from h_{20}. Specifically, working through the relatively straightforward (if tedious) algebra, the result is

$$\chi = a_{11}h_{20} - a_{21}h_{10} + \beta_+ y = \overline{\phi}\left(\frac{a_{11} - \gamma a_{21}}{1 + \gamma + \lambda}\right) + \beta\left(1 - \alpha_+ + \frac{a_{11}\lambda + a_{21}(1 + \lambda)\lambda}{1 + \lambda + \gamma}\right)y + \text{const}, \tag{2.43}$$

a plot of which appears in the western part of Fig. 2. The ratio of nonlinear to linear phase speeds in the western half basin is

$$\frac{-\chi_y}{-\beta_+} = 1 + \frac{a_{11} - a_{21}\gamma}{(1 + \lambda + \gamma)}\frac{\overline{\phi}_y}{\beta_+} + \frac{a_{11}\lambda + a_{21}\lambda(1 + \lambda)}{(1 - \alpha_+)(1 + \lambda + \gamma)}. \tag{2.44}$$

Some algebra shows that the coefficient in front of $\overline{\phi}_y$ is negative, which, given that $\overline{\phi}_y$ is negative in the north half of the subtropical gyre, makes that term positive. The last term on the right-hand side is also shown to be positive; thus, the westward wave movement is seen unambiguously to be accelerated relative to the linear rate.

It is of interest to compare the relative strengths of the ''advective'' effects in the eastern and western basins. Both may be written as $C_{\text{ph}} = -(\beta_+ + C_c)$, where C_c represents a correction to the linear westward phase speed. Equation (2.37) defines C_{ce}, the eastern correction. The total western phase speed may be written as

$$C_{\text{phw}} = \beta_+ + \frac{1}{q}\left[\frac{\beta_+\lambda\gamma}{1 + \gamma + \lambda} - \frac{\overline{\phi}_y}{2}\left(1 - \frac{q}{(1 + \lambda + \gamma)}\right)\right], \tag{2.45}$$

where q is the radical in (2.11) in the limit $H_3/H \to 1$.

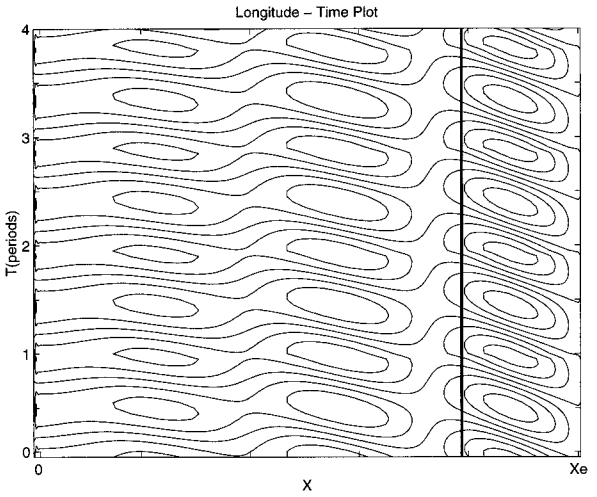

FIG. 3. Longitude–time plots from the analytical model. Longitude occurs on the x axis and time on the y axis. The associated latitude is in the northern half of the basin. The contours are of first-mode wave amplitude and the bold vertical line indicates the boundary between uniform potential vorticity to the west and the second layer stagnant zone to the east. Note the increase in wave phase speed as one moves west.

As the boundary between the nonuniform and uniform pool of second layer potential vorticity is crossed on a line of constant latitude, wave zonal phase speed changes discontinuously. To see this, the ratio of the corrections to the background β_+ phase speed is computed:

$$r = \frac{C_{\text{ce}}}{C_{\text{cw}}} = \left[-\frac{\beta\lambda}{\overline{\phi}^e_y}\left(\frac{1}{2} + \frac{q}{2(1 + \lambda + \gamma)}\right) + \frac{\overline{\phi}^w_y}{\overline{\phi}^e_y}\frac{1}{2}\left(1 - \frac{q}{(1 + \lambda + \gamma)}\right)\right]^{-1}, \tag{2.46}$$

where $\overline{\phi}^{e\,(w)}$ denotes a streamfunction value in the east (west) basin.

First, at the division between protected and unprotected regions, $\phi^e_y = \phi^w_y$ and $\beta\lambda/(-\overline{\phi}^e_y) > 1$ [see (2.20)], so generally $C_{\text{ce}} < C_{\text{cw}}$. Second, away from the boundary between protected and unprotected zones, $-\overline{\phi}^w_y > -\overline{\phi}^e_y$ generally; therefore, the ratio r will be larger than 1. This result suggests longitude–time plots at a given latitude should generally show waves in the west to be moving faster than waves in the east.

This is illustrated in Fig. 3, where such a plot produced from our analytical model appears. The parameter values for the plot are $\lambda = 4$, $\gamma = 3$, and $\beta = 1/6$. Contours of first-mode wave amplitude versus time are shown. The associated latitude is in the northern basin half. The bold vertical line occurs at the longitudinal boundary between the uniform q region and the stagnant region. Note the considerable increase in pattern speed as one moves across the basin from east to west.

f. Why?

In summary, we find in this model acceleration of baroclinic planetary waves, relative to their linear propagation rates, in the northern domain of a subtropical gyre; deceleration in the south; and differences between the eastern and western wave propagation rates. These results coincide with tendencies noted in CS. Even though we would therefore argue the model enjoys some success in explaining the observations, it must be confessed the results are thus far mathematical. It remains to shed light on why the curious result is obtained that the mean shear wave effect is opposite to the sense of the circulation. We examine here the reasons; for that purpose it is useful to revisit an appropriate form of (2.6).

For simplicity, consider the case only of a zonal mean flow and no forcing so that (2.6) may be rewritten as

$$h'_{1t} - (h_{10} + h_{20})_y h'_{1x} - (h'_1 + h'_2)_x(\beta - h_{10y}) = 0, \tag{2.47}$$

representing thermocline evolution due to mean state advection and wave advection of the mean state potential vorticity field.

Suppose, first, the mean thermocline takes a first baroclinic mode structure and the perturbations define only a first baroclinic mode structure. In that case, $h_{10} = \alpha_- \gamma h_{20}$ and $h'_1 = \alpha_- \gamma h'_2$, so (2.47) becomes

$$h'_{1t} - \underline{(1 + \alpha_-\gamma)h_{20y}h'_{1x}} - \beta_+ h'_{1x} + \underline{(1 + \alpha_-\gamma)h_{20y}h'_{1x}} = 0, \tag{2.48}$$

where properties of the quadratic equation associated with (2.11) have been used. Note that an exact cancellation occurs between the underlined mean state contributions, leaving behind the unperturbed propagation equation. This is in agreement with (2.32b); namely, the lack of a second mode in the mean flow predicts that the wave will move at the standard speed. The reason for this cancellation is that a first-mode thermocline can potentially enhance the vortex tube stretching associated with meridional fluid movements, but these same thermocline tilts imply a zonal shear and thus limit the extent to which the wave can tap into the enhanced source of stretching. The same reasoning for a two-layer model leads to the classic non-Doppler result. Here, we have shown in a multilayer model that if the thermocline has a first-mode structure only, stretching and advection cancel.

Now consider the opposite case; namely, the thermocline is exactly a second-mode thermocline, that is, $h_{10} = \alpha_+ \gamma h_{20}$. Then a first-mode wave perturbation obeys

$$h'_{1t} - \beta_+ h'_{1x} - (1 + \underline{\alpha_+\gamma})h_{20y}h'_{1x} + \underline{\alpha_+\gamma h_{20y}h'_{1x}} = 0, \tag{2.49}$$

where a term that ultimately forces a second-mode response has been neglected. The underlined terms cancel one another, indicating zonal advection still works against the meridional stretching caused by the meridional wave advection. Note, however, that the mean state advection is too weak to completely cancel the enhanced stretching, leaving a resultant that adds to the standard linear value.

If the thermocline is composed of a combination of the two modes, the first baroclinic mode parts cancel, leaving behind only that part due to the second mode. If the deeper shear is westward, the wave speed is augmented to the west; a similar result is discussed by Killworth et al. (1977). Conversely, if the deeper shear is eastward, the wave speed is retarded. This is also in agreement with (2.32b) and points out the sensitivity of the effect of shear to the sign of the second baroclinic mode. It is also essential to note that the wave speed is not determined solely by the potential vorticity field of the medium in which the wave resides; there is also an equally important advective contribution from the mean flow.

It was noted earlier that the wave–shear interaction induced a cyclonic correction to the wave characteristics for a subtropical gyre. This appeared in the result (2.36) and the positivity of (2.44). This result is now seen to reflect that the second-mode structure of the general circulation in this three-layer model has a positive amplitude [see (2.32b)]. Associated with this is the thermocline structure denoted in Fig. 4, which generates the required westward deep shear velocity and thus the wave enhancement.

g. Generalization to continuous stratification

It is of interest to consider the generalization of this analysis to a continuously stratified system; we will follow a procedure like that in Flierl (1978). The quasigeostrophic equation for a continuously stratified fluid is

$$\frac{\partial}{\partial z}\frac{f_0^2}{N^2}\frac{\partial}{\partial z}\psi_t + J\left(\psi, \frac{\partial}{\partial z}\frac{f_0^2}{N^2}\frac{\partial}{\partial z}\psi + \beta y\right) = 0, \tag{2.50a}$$

where we work in dimensional coordinates and neglect relative vorticity due to our interests in scales that are long compared to the deformation radius. The quantity ψ is the streamfunction and N denotes the (known) buoyancy frequency. To (2.50a), we append the boundary conditions:

$$\frac{f_0}{N^2}(\psi_{zt} + J(\psi, \psi_z)) = -w_e \quad \text{at } z = 0 \tag{2.50b}$$

$$\frac{f_0}{N^2}(\psi_{zt} + J(\psi, \psi_z)) = 0 \quad \text{at } z = -H, \tag{2.50c}$$

where, for convenience, we ignore bottom friction.

The linear problem based on (2.50) yields the Sturm–Liouville problem for the vertical structure functions F_n:

$$\frac{\partial}{\partial z}\frac{f_0^2}{N^2}\frac{\partial}{\partial z}F_n + \lambda_n^2 F_n = 0, \tag{2.51}$$

which satisfy the boundary conditions:

$$F_{nz} = 0 \quad \text{at} \quad z = 0, -H \tag{2.52a}$$

and are normalized according to

$$\frac{1}{H}\int_{-H}^{0} F_i F_j \, dz = \delta_{ij} = \begin{cases} 1, & i = j \qquad (2.52b) \\ 0, & i \neq j. \qquad (2.52c) \end{cases}$$

The streamfunction ψ is now written as $\psi_0 + \psi'$, where ψ_0 solves the steady form of (2.50). Projecting the resulting predictive equation for ψ' on F_0 eventually yields the equation

$$\beta\alpha_{0x} = \frac{f_0}{H}w_e', \tag{2.53}$$

where

$$\alpha_0 = \frac{1}{H}\int_{-H}^{0} F_0\psi' \, dz \tag{2.54}$$

represents the horizontal structure of the barotropic wave mode. The equation for the first baroclinic wave mode is

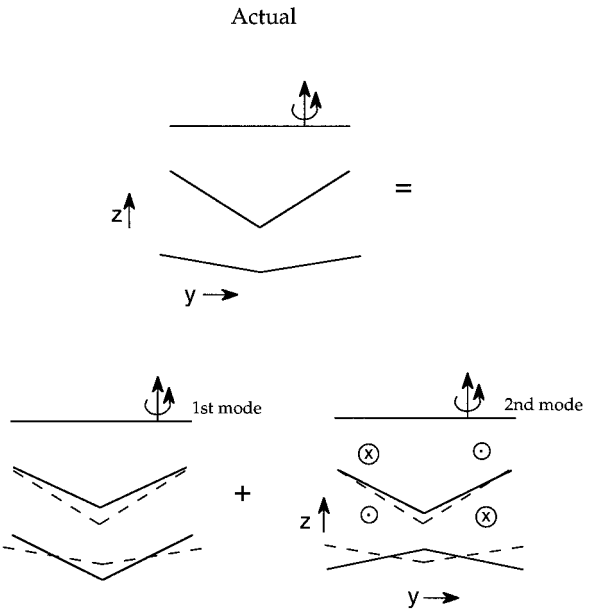

FIG. 4. Wave propagation schematic. The thermocline appears in the top panel as the solid lines, and in the following panels as the dotted lines. The thermocline structure is in general composed of a first- and second-mode contribution and this appears schematically in the diagram where the first and second modes appear as the solid lines in the bottom two panels. The general circulation structure tends to be more surface confined than the first mode, which requires the presence of a positive second-mode amplitude, as shown. This in turn possesses the westward deep shear required for the westward acceleration of planetary waves in the northern half basin.

$$-\lambda_1^2\alpha_{1t} + \sum_i J\left[\alpha_i, \frac{1}{H}\int_{-H}^{0}\left(F_i F_1 \frac{\partial}{\partial z}\frac{f_0^2}{N^2}\frac{\partial}{\partial z}\psi_0\right)dz - \frac{1}{H}\int_{H}^{0} F_i\frac{\partial}{\partial z}\frac{f_0^2}{N^2}\frac{\partial}{\partial z}(F_1\psi_0)\, dz\right]$$
$$+ \frac{1}{H}\int_{-H}^{0} F_1 J\left(\psi', \frac{\partial}{\partial z}\frac{f_0^2}{N^2}\frac{\partial}{\partial z}\psi'\right)dz + \beta\alpha_{1x} = \frac{-F_1(0)f_0 w_e'}{H}. \tag{2.55}$$

In keeping with the earlier analysis, we neglect all wave–wave nonlinear interactions, anticipating their removal by an averaging method. Similarly, we retain only the first-mode interactions with the general circulation. These two constraints reduce to the neglect of terms quadratic in ψ' in (2.55) and to the restriction that $i = 0$ and 1 only in the summations. Now the projection of the mean flow, ψ_0, on the modes (2.51) is introduced as

$$\mu_K = \frac{1}{H}\int_{-H}^{0} F_K\psi_0 \, dz,$$

and the mode–mode interaction coefficient

$$\xi_{ijk} = \frac{1}{H}\int_{-H}^{0} F_i F_j F_k \, dz$$

is defined. After several integrations by parts, (2.55) reduces to

$$\alpha_{1t} + J\left[\sum_K\left(\frac{(\lambda_1^2 - \lambda_K^2)}{\lambda_1^2}\right)\xi_{11K}\mu_K + \frac{\beta y}{\lambda_1^2}, \alpha_1\right] = \frac{-F_1(0)f_0 w_e'}{H\lambda_1^2} - J(\mu_1, \alpha_0). \tag{2.56}$$

The above is analogous to (2.32b), as can be verified upon inspection. In particular, note that the projection of the mean flow on the first baroclinic mode makes no contribution to the evolution of α_1. This is seen upon substituting $K = 1$ in (2.56) and is entirely consistent with the earlier three-layer analysis. Note also that the mean structure effect on wave propagation acts like an advective term, also in agreement with (2.32b). Finally, the alteration of the mean potential vorticity field as an effect on the production of waves, measured by the last term in (2.56), vanishes as $H \to \infty$ [see (2.53)].

The fundamental structure of (2.56) is much like that of the three-layer result, which leads to some notable conclusions. Most important is that the minimal model capable of capturing the essential dynamics of wave propagation in a moving ocean is a three-layer model. The progression from the simplest models onward adds new effects in an orderly sequence: a reduced gravity model yields the standard linear result if the ocean is at rest; if it is moving, propagation is modified to include the local variation in deformation radius; a two-layer model illustrates the Doppler shift by a depth-averaged flow and a three-layer model exhibits shear interaction. Going on to an infinite number of layers (via the use of continuous stratification) is then seen, at least in QG dynamics, not to add new physics, although quantitative predictions improve.

We found in the three-layer model that the shear term augmented the linear speed. Ultimately, this was seen to reflect the presence of a positive projection of the general circulation on the second baroclinic mode. It is interesting to ask if a similar sign-definite result can be obtained here. For a continuously stratified ocean, the mean field is governed by

$$\frac{\partial}{\partial z}\frac{f_0^2}{N^2}\frac{\partial}{\partial z}\psi_0 + \beta y = q(z, \psi), \tag{2.57}$$

where $q(z, \psi)$ is the potential vorticity field. According to Rhines and Young theory, $q(z, \psi) = \beta y_N > \beta y$ for the moving parts of layers in a subtropical gyre. If it is accepted that $q(z) \geq \beta y$ routinely on subsurface isopycnals, then the positivity of the amplitude of the general circulation modes can be supported by multiplying (2.57) by F_n and integrating. The result after two integrations by parts is

$$F_N\frac{f_0^2}{N^2}\psi_{0z}\Big|_{-H}^{0} - \lambda_n^2 H\mu_n = \int_{-H}^{0} F_n\delta q\, dz, \tag{2.58}$$

where $\delta q = q(\psi, z) - \beta y > 0$. A further integration of (2.57) and some algebra leads eventually to

$$\mu_n = \frac{1}{\lambda_n^2 H}\int_{-H}^{0}[F_n(0) - F_n(z)]\delta q\, dz, \tag{2.59}$$

where the general circulation result $\psi_{0z}(-H) = 0$ has been used. While not always the case, it is often true that $F_n(0) - F_n(z) > 0$ for oceanic modes [see Pedlosky (1979) for examples]. Under the assumption that σq is positive, one sees $\mu_n > 0$ from (2.59). The sense of the mean flow effect on first-mode wave propagation from (2.56) then follows $(\lambda_1^2 - \lambda_k^2)$, which is unambiguously negative from Sturm–Liousville theory. The effect of a subtropical gyre on waves in a continuously stratified fluid is seen to be qualitatively like that in a three-layer fluid and generally to induce a cyclonic correction to wave phase speeds.

The tendency for $\mu_n > 0$ noted in (2.59) reflects that the general circulation is largely surface intensified. The first mode reflects in-phase movements over the whole of the thermocline. The addition of higher modes results in a mean structure confined nearer to the surface.

3. Verification of the wave model

The analysis leading to (2.32) is formal and depends upon a number of assumptions. Most importantly, it is valid in the limit of a large-scale separation between the first and second baroclinic modes. In addition, the important assumption was made that the main attributes of the modes found in linear theory persist in the nonlinear problem. In this section, the theory is tested by comparing solutions obtained from it to numerical solutions of the full system.

For simplicity, we work with the equations valid in the limit of a very deep third layer. The approximate equation governing the first mode is then (2.33), while the complete equations are obtained from (2.6) and (2.7) and are

$$h_{1t} - J(h_1 + h_2, \beta y - h_1) = -\beta w_e + K\nabla^2 h_1 \tag{3.1}$$

$$(h_1 - \gamma h_2)_t + J(h_2, \beta\lambda y + h_1 - \gamma h_2) = K\nabla^2(h_1 - \gamma h_2). \tag{3.2}$$

Note that thickness diffusion, necessary for numerical stability, has been added to the above. The numerical methods employed here are standard. A finite difference model was built; Jacobians were computed using the fourth-order accurate, enstrophy and energy conserving scheme of Arakawa (1966); a leapfrog method with infrequent averaging of adjacent time levels was used for time stepping; and the diffusive terms were lagged to ensure stability. The form of the Ekman pumping in (3.1) was

$$w_e = (a_0 + a_1\cos(\omega t))\sin(\pi y), \tag{3.3}$$

for $y = (-1, 0)$; that is, we studied a single subtropical gyre circulation only. In all experiments shown here, the parameters λ, β, and γ were set to 1.

We have found (3.1) and (3.2) surprisingly demanding to solve because of the presence of the short second-mode wave. At a resolution of $(251)^2$ points, the dissipation necessary to stabilize the model was found to damp the second-mode amplitude within a few wavelengths of the western boundary. To obtain a relatively undamped solution, it was necessary in our best experiments to either go to $(501)^2$ resolution, and employ diffusivities of roughly 10^{-4}, or to employ higher-order diffusivities. Dimensionally, the values of 10^{-4} correspond to diffusivities of $0(1\ \mathrm{m^2\ s^{-1}})$, that is, well below (by a factor of 20–200) what is routinely used in ocean simulation. Even then, it was necessary to conduct separate wave–wave and wave–mean flow experiments to study the impact of the smallest scales.

a. Wave–wave experiments

We first studied (3.1) and (3.2) with $a_1 = 1$ and $a_0 = 0$ in (3.3). That is, we initially excluded a time-mean

Ekman pumping but allowed vigorous, variable pumping. Our analytic theory argues this will result in the lack of a mean circulation, and thus that the first-mode wave will to leading order be indistinguishable from a purely linear first-mode wave [viz., (2.33) with $h_{10} = h_{20} = 0$].

We show in Figs. 5a,b plots of h_1 and h_2 as computed by (3.1)–(3.2) from which it is clear that the dominant behavior resembles a sequence of propagating large-scale waves, on top of which is superimposed variability of a considerably smaller scale. The longer spatial scale represents the first baroclinic mode, and the shorter scale the second baroclinic mode. Thus, clearly nonlinear interactions in this experiment are fully active.

The present theory governs h_+ rather than h_1 and h_2; therefore, we constructed h_+ from these numerical results by straightforward manipulation. The result appears in Fig. 5c. Note the lack of a short spatial scale here, consistent with expected first-mode structure. A comparison of the numerical h_+ and the ''theoretical'' h_+ is given in Figs. 5c,d. Here, the theoretically generated h_+ was computed as a numerical solution to (2.33). Comparisons of it with the analytical solution to (2.33) showed them to be virtually the same. The point of Figs. 5c and 5d is that the theoretical, and in this case linear, wave constitutes an excellent representation of the full solution. This is further reinforced in Fig. 5e, where a more direct comparison is given. Here, first-mode amplitudes at the middle latitude of the basin from both the numerical and theoretical predictions are plotted versus longitude. The latter has been offset vertically by a small amount in order that the two curves could be seen. (Otherwise, the curves, being virtually coincident, reduce to one line.) In any case, it is noted that the agreement is excellent. We conclude from this experiment that, in agreement with the structure of our theory, nonlinear exchanges among the baroclinic wave modes are a relatively weak effect. The analysis capitalizes on a scale separation and, thus, an impedance to interaction due to a structural mismatch between modes to eliminate these effects. The sense of the numerical experiments, as seen in Fig. 5, supports this.

b. Wave–mean flow experiments

We now discuss experiments in which both a_0 and a_1 were set to unity. This alteration made the requirement on the diffusivity coefficient K for numerical stability somewhat more strict. It was necessary at a $(501)^2$ resolution to employ a diffusivity of 10^{-3} to achieve stable solutions. Mostly this was due to the presence of the western boundary current, although the southwest corner of the domain also tended to develop structures at very small, but resolved, scales. The reasons for the latter behavior seem to be related to a reversal with depth of potential vorticity gradient; that is, the surface potential vorticity decreases in value to the north, while the second layer is dominated by β. The southwest quadrant is thus potentially baroclinically unstable, and this results in the production of small scales. Experiments with different K showed that these small scales could be controlled by diffusion; thus, we proceeded with experiments using $K = 10^{-3}$ (dimensionally $K \sim 10$ m^2 s^{-1}). We note, however, that these diffusion levels are sufficient to damp structure on the scale of the second-mode waves within a few wavelengths of the eastern boundary. We appeal to the results of the previous wave–wave interaction section to argue that this effective loss of the short second-mode scale is not a serious flaw.

The model solving (3.1)–(3.2) was started from a state of rest and integrated to nondimensional time of 3. First-mode waves require a time of 0.4 to cross the basin; thus, this represents roughly 7 first-mode crossings. In comparison, the frequency of the variable forcing ω was 20π, requiring a time of 0.1 to complete an oscillation. Then, the results from 3.0 to 3.1, that is, from one complete forcing cycle, were saved at 20 equally spaced time intervals. These results were then averaged to obtain an estimate of the mean fields h_{10} and h_{20}, with the latter estimates being used in three ways.

First, they were subtracted from the complete fields to produce maps of the variability of h_1 and h_2 (called h_{11} and h_{21} in the notation of the theory). Maps of h_+ were then constructed from these residual maps according to the formula $h_+ = h_1 - \alpha_+ \gamma h_{21}$.

Second, the computed h_{10} and h_{20} fields were compared to both theoretical mean fields and numerically generated mean fields. The comparison with the theoretical fields demonstrated that the gross features of the h_{10} and h_{20} fields were in basic agreement with the analytical steady theory. The numerical results, however, did not develop the pool of uniform potential vorticity predicted by the theory, although the location of the transition point predicted in (2.37) was well modeled. This is due to the lack of relative vorticity in the large-scale quasigeostrophic model. Thus we also compared the numerically generated h_{10} and h_{20} fields with mean numerical fields generated in the absence of variable wind forcing (i.e., $a_1 = 0$), in principle to clarify the generation of mean flow by wave–mean flow interaction. In all cases, the comparison was excellent, demonstrating the mean fields were largely independent of the variability.

Third, the estimated means were used to generate a theoretically expected field of h_+ by integration of (2.33). Comparisons of the numerical and theoretical results appear in Figs. 6a–d. These fields occur every quarter of a cycle and, except for the southwestern quadrant, show that the analytical and numerical solutions for first-mode baroclinic wave propagation are, to an excellent degree, in agreement.

In particular, note that the wave crests are very structured in the north–south direction. Near the eastern boundary, the waves are nearly meridionally aligned. This occurs for two reasons: First, in the quasigeo-

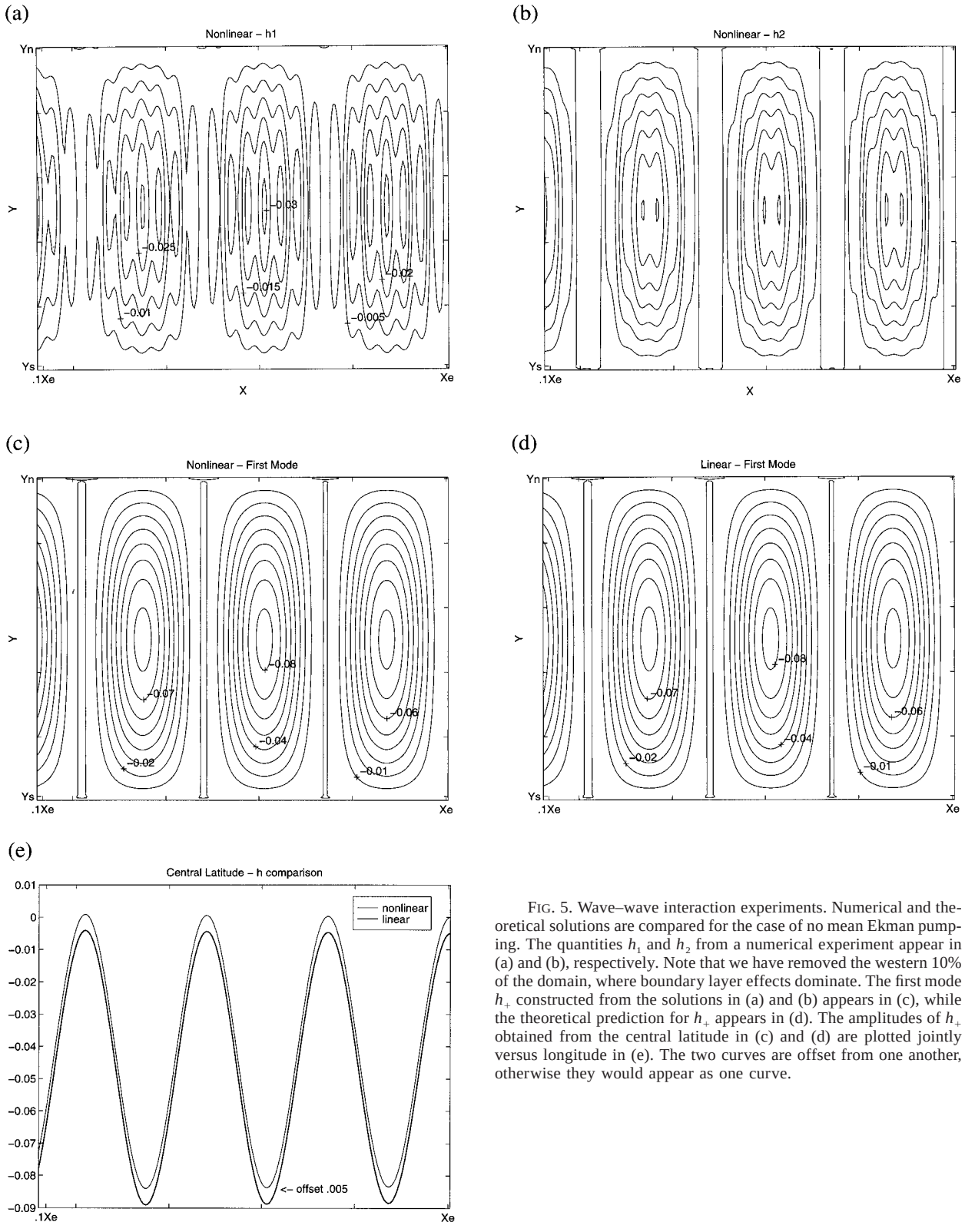

FIG. 5. Wave–wave interaction experiments. Numerical and theoretical solutions are compared for the case of no mean Ekman pumping. The quantities h_1 and h_2 from a numerical experiment appear in (a) and (b), respectively. Note that we have removed the western 10% of the domain, where boundary layer effects dominate. The first mode h_+ constructed from the solutions in (a) and (b) appears in (c), while the theoretical prediction for h_+ appears in (d). The amplitudes of h_+ obtained from the central latitude in (c) and (d) are plotted jointly versus longitude in (e). The two curves are offset from one another, otherwise they would appear as one curve.

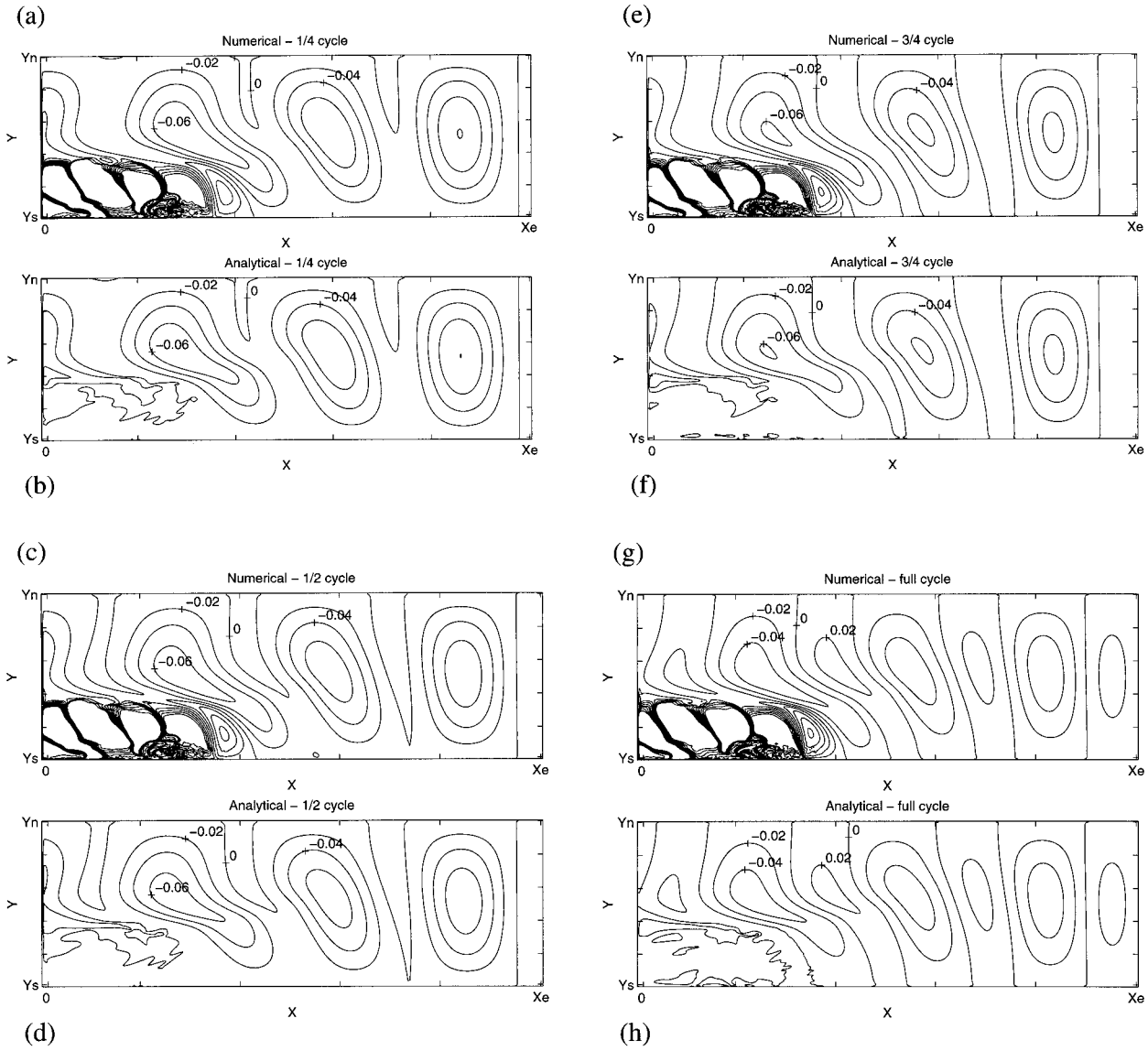

FIG. 6. Wave–mean flow experiments. Numerical and theoretical solutions for first-mode wave propagation are compared. The numerical solutions at four equally spaced intervals from a forcing cycle appear in (a), (c), (e), and (g). The results from (3.23) at the same points in the cycle appear in (b), (d), (f), and (h). Aside from the southwest quadrant, where the complete model develops small scales, the agreement is excellent.

strophic approximation, there is no dependence of wave phase speed on latitude; thus, near x_e, the wave fronts are strictly meridional. Second, the strength of the mean flow effect on the zonal wave propagation grows with distance from x_e. This was seen in section 2 when the analytic solution of (2.33) was discussed. Thus, it is only farther in the basin that clear distinctions between the northern and southern half circulations, in regards to their wave characteristics, arise. Then it is seen that the northern half basin leads the southern half basin. This reflects acceleration (deceleration) of the waves in the northern (southern) parts of the circulation. Finally, the apparent production of smaller scales, clearest in Fig. 5 (panels b, g, and h) reflects the interaction of the propagating and standing wave signals in (2.42b).

We interpret the above comparison as evidence supporting the aptness of our theoretical development.

4. Summary and discussion

This paper was motivated by the recent analysis of Chelton and Schlax (1996) of TOPEX/Poseidon altimetric data, which shows sea surface anomalies moving at speeds too fast to be convincingly ascribed to linear planetary waves. We therefore have reconsidered the general problem of low-mode planetary wave dynamics.

Various considerations led us to examine the effects of a mean shear.

The results are encouraging with respect to providing an explanation for the observations and they are also surprising given previous results. Regarding the first point, our analysis shows westward wave phase speeds will be accelerated in the northern half of a subtropical gyre, and decelerated in the southern half. Also, under reasonable oceanic conditions, the acceleration rates relative to the linear speeds are easily a factor of 2. This relative acceleration is certainly in keeping with CS; it is also curious that the deceleration predicted at lower latitudes is also seen in the CS results. This latter comparison is much less certain, and perhaps fortuitous.

The results are surprising in that they run entirely counter to what one would expect given the sense of the circulation. Specifically, the subtropical gyres are anticyclonic, yet the nature of the wave correction is cyclonic. As a partial explanation of this, it was pointed out in the analysis of the three-layer system that an anticyclonic correction of the wave phase speeds was indeed driven by a Doppler-like effect. The source of this correction has to be found in the depth-averaged flow and provided an effect inversely proportional to the fluid depth. The impact of this correction was studied in Dewar (1989) and was subsequently removed from the present problem by considering a very deep third layer. The remaining corrections could thus be unambiguously connected to the wave interaction with the mean-shear structure. There it was found that the projection of the mean thermocline of a subtropical gyre on the second baroclinic mode induced a cyclonic correction to the wave speed. Both advection and stretching effects from the tilted isopycnals contribute importantly. Subtropical gyres tend to have a positive second-mode structure, which results in the acceleration of the waves to the west in the northern half basin. The effect reverses in sign for a negative second baroclinic mode structure. The lack of an effect from the first baroclinic mode represents the presence of the non-Doppler effect in this model. The above results, found in an analysis of a QG model, have also been numerically tested. The comparisons support the accuracy of the theory.

Extensions to continuously stratified systems were discussed and we are prepared to argue that much of the salient dynamics of this model also occur in the planetary geostrophic equations (see the appendix). The key point of the analysis on which this is based is again the assumption of a scale separation between the mean circulation and the waves, with the first-mode wave comparatively much larger than the second. Exploiting these scale distinctions leads to equations like (2.33). The $O(1)$ stratification variations found in the planetary geostrophic equations modify the strength of the interactions, but much of the dynamics are still the wave–mean flow interactions found in the QG model.

In summary, of the many possible explanations offered for the CS observations, that is, mean flow effects, topography, forcing, and atmosphere–ocean coupling, our analysis applies to the first and suggests it is of the right size and sign to play a role in ocean wave propagation. This is in agreement with the findings of Killworth et al. (1997), who provide considerable quantitative support to this idea though their analysis of existing hydrographic data for the world's oceans. Our approaches are, however, quite distinct and mutually complementary.

Acknowledgments. WKD is supported by NASA Grant NAGW-4883 and NSF Grants OCE-9401977 and OCE-9617728. This work was initiated during discussions between the authors during a visit by WKD to Oregon State University. The hospitality of that institution is gratefully acknowledged. Sheila Heseltine prepared the manuscript. Jane Jimeian assisted with computational issues.

APPENDIX

Extension to Planetary Geostrophic Equations

We consider here a three-layer planetary geostrophic model in the limit of a deep third layer. The geometry is similar to Fig. 1, except that the quantity h_1 (h_2) now refers to the total layer thickness, rather than to the small variation of a given interface about a uniform depth. After algebra like that in section 2, the dimensional equations governing the two layers are

$$h_{1t} + \frac{g_2'}{f}J(h_2, h_1) - \frac{\beta}{f^2}g_2'h_1(h_1 + h_2)_x - \frac{\beta g_1'}{f^2}h_1h_{1x} = -w_e \tag{A.1}$$

$$h_{2t} + \frac{g_2'}{f}J(h_1, h_2) - \frac{\beta}{f^2}g_2'h_2(h_1 + h_2)_x = 0, \tag{A.2}$$

where $f = f(y)$.

Now, we consider the linearized problem, in which $h_1 = h_{10} + h_{11}$, $h_2 = h_{20} + h_{21}$, and (h_{10}, h_{20}) are considered constant. The perturbations h_{11} and h_{21} are thus governed by

$$h_{11t} - \frac{\beta}{f^2}g_2'h_{10}(h_{11} + h_{21})_x - \frac{\beta g_1'}{f^2}h_{10}h_{11x} = -w_e' \tag{A.3}$$

$$h_{21t} - \frac{\beta}{f^2}g_2'h_{20}(h_{11} + h_{21})_x = 0. \tag{A.4}$$

The normal-mode equations for $h_{11} + \alpha_\pm h_{21}$ requires

$$\alpha_\pm = \frac{(g_2'h_{20} - (g_1' + g_2')h_{10}) \pm [(g_2'h_{20} - (g_1' + g_2')h_{10})^2 + 4g_2'^2h_{20}h_{10}]^{1/2}}{2g_2'h_{20}}. \tag{A.5}$$

The linear problem predicts a length-scale separation like that found in the QG problem; that is, the waves are considerably shorter than the basin scale. We shall assume this persists into the nonlinear problem. With this, it is clear that in the presence of variable h_{10} and h_{20} (associated with the general circulation) $\alpha_\pm$ can be used locally to define the first and second modes. Note that this implies [representing the basin scale with coordinates (x, y)] $\alpha_\pm = \alpha_\pm(x, y)$.

Returning now to (A.1) and (A.2), we write $h_1 = h_{10} + h_{11}$, $h_2 = h_{20} + h_{21}$, where h_{10} and h_{20} come from ventilated thermocline theory and are thus known. The equations governing h_{11} and h_{12} are

$$h_{11t} + \frac{g_2'}{f}(J(h_{20}, h_{11}) + J(h_{21}, h_{10}) + J(h_{21}, h_{11})) - \frac{\beta g_2'}{f^2}h_{10}(h_{11} + h_{21})_x - \frac{\beta}{f^2}g_2'h_{11}(h_{10} + h_{20})_x$$
$$- \frac{\beta}{f^2}g_2'h_{11}(h_{11} + h_{21})_x - \frac{\beta g_1'}{f^2}h_{11}h_{10x} - \frac{\beta g_1'}{f^2}h_{10}h_{11x} = -w_e' \quad \text{(A.6)}$$

$$h_{21t} + \frac{g_2'}{f}\{J(h_{10}, h_{21}) + J(h_{11}, h_{20}) + J(h_{11}, h_{21})\} - \frac{\beta g_2'}{f^2}h_{20}(h_{11} + h_{21})_x$$
$$- \frac{\beta g_2'}{f^2}(h_{21}(h_{10} + h_{20})_x + h_{21}(h_{11} + h_{21})_x) = 0. \quad \text{(A.7)}$$

The equation governing the first mode h_+ is formed from (A.6) and (A.7) by straightforward manipulation.

Recalling the scaling in section 2, it is anticipated that the wave amplitudes will be small, but their derivatives will not be. This reflects that the waves occur on small scales and that their relative derivatives need to occur on these scales. This recognition is at the heart of the earlier multiple scales analysis, which could be repeated here. Rather than that, however, we will simply quote the result. Along the way, we will have rewritten h_{11} and h_{21} using h_+ and h_- (and the familiar a_{ij} notation), neglected undifferentiated h_+ and h_-, and also differentiated h terms. The latter are dropped following the ''method of averaging'' approach. The result is

$$h_{+t} - \frac{\beta_+}{f^2}h_{+x} + \frac{g_2'}{f}(1 - \alpha_+)$$
$$\times (a_{11}J(h_{20}, h_+) - a_{21}J(h_{10}, h_+)) = -w_e', \quad \text{(A.8)}$$

where

$$\beta_+ = \beta(g_2'h_{10} + g_1'h_{10} + g_2'\alpha_+h_{20}) \quad \text{(A.9)}$$

and α_+ is given by (A.5).

Equation (A.8) should be compared to (2.33) in the text. Note that the differences are mainly quantitative and reflect that, for example, the deformation radius is subject to local definition. The form of (A.8) compares favorably with (2.33), suggesting that the QG analysis has captured much of the basic dynamics.

REFERENCES

Arakawa, A., 1966: Computational design for long-term numerical integration of the equations of fluid motion: Two-dimensional incompressible flow. Part I. *J. Comput. Physics,* **1,** 119–143.

Barnier, B., 1986: Investigation of the seasonal variability of the wind stress curl over the North Atlantic Ocean by means of E.O.F. analysis. *J. Geophys. Res.,* **91,** 863–868.

——, 1988: Numerical study on the influence of the mid-Atlantic ridge on nonlinear first-mode baroclinic Rossby waves generated by seasonal winds. *J. Phys. Oceanogr.,* **18,** 417–433.

Bender, C., and S. Orszag, 1978: *Advanced Mathematical Methods for Scientists and Engineers.* McGraw-Hill, 593 pp.

Bernstein, R. L., and W. B. White, 1974: Time and length scales of baroclinic eddies in the central North Pacific Ocean. *J. Phys. Oceanogr.,* **4,** 613–624.

——, and ——, 1977: Zonal variability in the distribution of eddy energy in the mid-latitude North Pacific Ocean. *J. Phys. Oceanogr.,* **7,** 123–126.

Chang, P., and S. Philander, 1989: Rossby wave packets in baroclinic mean currents. *Deep-Sea Res.,* **36,** 17–37.

Chelton, D. B., and M. G. Schlax, 1996: Global observations of oceanic Rossby waves. *Science,* **272,** 234–238.

Cummins, P. F., L. A. Mysak, and K. Hamilton, 1986: Wind-stress curl generation of annual Rossby waves in the North Pacific. *J. Phys. Oceanogr.,* **16,** 1179–1189.

de Szoeke, R. A., and D. B. Chelton, 1998: The modification of long planetary waves by homogeneous potential vorticity layers. *J. Phys. Oceanogr.,* in press.

Dewar, W. K., 1989: A nonlinear, time-dependent thermocline theory. *J. Mar. Res.,* **47,** 1–31.

Emery, W., and L. Magaard, 1976: Baroclinic Rossby waves as inferred from temperature fluctuation in the eastern Pacific. *J. Mar. Res.,* **34,** 365–385.

Flierl, G. R., 1978: Models of vertical structure and the calibration of two-layer models. *Dyn. Atmos. Oceans,* **2,** 341–381.

Gill, A., 1982: *Atmosphere–Ocean Dynamics.* Academic Press, 662 pp.

Held, I. M., 1983: Stationary and quasi-stationary eddies in the extratropical troposphere: Theory. *Large-Scale Dynamical Processes in the Atmosphere,* P. J. Hoskins and R. P. Pierce, Eds., Academic Press, 127–168.

Hermann, P., and W. Kraus, 1989: Generation and propagation of annual Rossby waves in the North Atlantic. *J. Phys. Oceanogr.,* **19,** 727–744.

Kang, Q., and L. Magaard, 1980: Annual baroclinic Rossby waves in the central North Pacific. *J. Phys. Oceanogr.,* **10,** 1156–1167.
Keefer, T., 1985: Ventilation of the world's oceans: Maps of the potential vorticity field. *J. Phys. Oceanogr.,* **15,** 509–523.
Kessler, W., 1990: Observations of long Rossby waves in the northern tropical Pacific. *J. Geophys. Res.,* **95,** 5183–5218.
Killworth, P. D., R. S. de Szoeke, and D. B. Chelton, 1997: The speed of observed and theoretical long extratropical planetary waves. *J. Phys. Oceanogr.,* **27,** 1946–1966.
Kraus, W., and C. Wuebber, 1982: Response of the North Atlantic to annual wind variations along the eastern coast. *Deep-Sea Res.,* **29,** 851–864.
Leetmaa, A., and A. Bunker, 1978: Updated charts of the mean annual wind stress, convergence in the Ekman layers, and Sverdrup transports in the North Atlantic. *J. Mar. Res.,* **36,** 311–322.
Lippert, A., and P. Kase, 1985: Stochastic wind forcing of baroclinic Rossby waves in the presence of a meridional boundary. *J. Phys. Oceanogr.,* **15,** 184–194.
McDowell, S., P. Rhines, and T. Keffer, 1982: North Atlantic potential vorticity and its relation to the general circulation. *J. Phys. Oceanogr.,* **12,** 1417–1436.
Meng, H., and W. K. Dewar, 1995: The propagation of submesoscale coherent vortices. *J. Phys. Oceanogr.,* **25,** 1745–1770.
Mysak, L. A., 1983: Generation of annual Rossby waves in the North Pacific. *J. Phys. Oceanogr.,* **13,** 1910–1923.
Pedlosky, J., 1979: *Geophysical Fluid Dynamics.* Springer-Verlag, 624 pp.
——, 1984: The equations for geostrophic motion in the ocean. *J. Phys. Oceanogr.,* **14,** 448–455.
Price, J., and L. Magaard, 1980: Rossby wave analysis of the baroclinic potential energy in the upper 500 meters of the North Pacific. *J. Mar. Res.,* **38,** 249–264.
——, and ——, 1983: Rossby wave analysis of subsurface temperature fluctuations along the Honolulu–San Francisco great circle. *J. Phys. Oceanogr.,* **13,** 258–268.
——, and ——, 1986: Interannual baroclinic Rossby waves in the midlatitude North Atlantic. *J. Phys. Oceanogr.,* **16,** 2061–2070.
Rhines, P., and W. Young, 1982: A theory of the wind-driven circulation I. Mid-ocean gyres. *J. Mar. Res.,* **40**(Suppl.), 559–596.
Sturges, W., and B. G. Hong, 1995: Wind forcing of the Atlantic thermocline along 32°N and low frequencies. *J. Phys. Oceanogr.,* **25,** 1706–1715.
——, B. G. Hong, A. Clarke, and X. Liu, 1998: Decadal wind forcing of the North Atlantic subtropical gyre. *J. Phys. Oceanogr.,* **28,** 659–668.
Talley, L., 1988: Potential vorticity distribution in the North Pacific. *J. Phys. Oceanogr.,* **18,** 89–106.
White, W., 1977: Annual forcing of baroclinic long waves in the tropical North Pacific. *J. Phys. Oceanogr.,* **7,** 50–61.
——, and J. F. T. Saur, 1981: A source of annual baroclinic waves in the eastern subtropical North Pacific. *J. Phys. Oceanogr.,* **11,** 1452–1462.