Initially Forced Long Planetary Waves in the Presence of Nonzonal Mean Flow

IVANA CEROVEČKI* AND ROLAND A. DE SZOEKE

College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon

ABSTRACT

The purpose of this paper is to understand how long planetary waves evolve when propagating in a subtropical gyre. The steady flow of a wind-driven vertically sheared model subtropical gyre is perturbed by Ekman pumping that is localized within a region of finite lateral extent and oscillates periodically at about the annual frequency after sudden initiation. Both the background flow and the infinitesimal perturbations are solutions of a 2½-layer model. The region of forcing is located in the eastern part of the gyre where the steady flow is confined to the uppermost layer (shadow zone). The lateral scales of the forcing and of the response are supposed to be small enough with respect to the overall gyre scale that the background flow may be idealized as horizontally uniform, yet large enough (greater than the baroclinic Rossby radii) that the long-wave approximation may be made. The latter approximation limits the length of time over which the solutions remain valid. The solutions consist of (i) a forced response oscillating at the forcing frequency in which both stable (real) and zonally growing (complex) meridional wavenumbers are excited plus (ii) a localized transient structure that grows as it propagates away from the region of forcing. Application of the method of stationary phase provides analytical solutions that permit clear separation of the directly forced part of the solution and the transient as well as estimation of the temporal growth rate of the transient, which proves to be convectively unstable. The solutions presented here are relevant to understanding the instability of periodic (including annual period) perturbations of oceanic subtropical gyres on scales larger than the baroclinic Rossby radii of deformation.

1. Introduction

a. Background and principal results

Chelton et al. (2004) analyzed four-year averages of 25-km-resolution measurements of near-surface wind speed and direction and showed that they reveal the surprisingly persistent small-scale features in ocean winds. Superimposed on these persistent wind patterns are temporally varying perturbations with magnitudes comparable to the mean winds. The strongest of these perturbations is the annual cycle. Chelton and Schlax (1996) correspondingly documented energetic annual-period Rossby wave sea surface elevation perturbations.

Motivated by these observations, the goal of this paper is to explain in detail how perturbations of a particular background subtropical gyre flow develop in space and in time when forced by winds characterized by scales much smaller than the gyre scale but much larger than the baroclinic Rossby radii of deformation that are started up at some initial instant and are thereafter periodic in time.

The solutions found here thus consist of a transient component and a component harmonic in time at the forcing frequency. Within the confines of the approximations and restrictions noted below, it is found that both the initial transient and the co-oscillating periodic response to spatially localized, suddenly initiated periodic Ekman pumping may be completely described by approximate analytical techniques.

The most important results are as follows. The perturbations excited by a patch of Ekman pumping that is localized within the eastern region of a subtropical gyre and oscillates periodically after being suddenly initiated

* Current affiliation: Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Corresponding author address: Ivana Cerovečki, 54-1721 Dept. of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139-4307.
E-mail: ivana@rossby.mit.edu

consist (i) of a directly forced solution that includes a well-defined beam of waves oscillating at the forcing frequency and radiating westward away from the forcing region without spatial growth plus a further harmonically oscillating part growing spatially toward the southwest as well as (ii) of a localized transient structure that grows as it propagates away from the region of forcing. Analytical approximate solutions permit clear separation of the forced and the transient solution as well as estimation of the growth rate of the transient, which proves to be convectively rather than absolutely unstable (Huerre and Monkewitz 1990), meaning that growth occurs only for certain ranges of x/t, y/t as $|x|$, $|y|$, $|t| \to \infty$, rather than for all x/t, $/y/t$.

b. Previous studies

Chelton and Schlax (1996) suggested that purely periodic (annual period) wind-driven long planetary waves can become energized in the western part of ocean basins. Explanations of this in the literature involve either (i) modification by midocean ridge topography of long waves radiated from the eastern boundary (Barnier 1988; Tailleux and McWilliams 2002) or else (ii) baroclinic instability (Liu 1999a; Dewar and Huang 2001; Kubokawa and Nagakura 2002). This paper further explores (ii) the instability. Previous studies in the literature that analyzed stability of long baroclinic waves in the subtropical gyre are as follows. Liu (1999a) found a linearized planetary wave instability in the southwestern part of a subtropical gyre that was associated with coupling, by the background subtropical gyre flow, of the fast and slow modes of a 2½ layer QG model. Liu (1999a) perturbed the steady flow of an eddy-resolving 3-layer QG ocean model with spatially uniform 3-yr period Ekman pumping and correspondingly found the 3-yr period ocean response to be "almost unchanged in longitude until it enters the southwestern part of the subtropical gyre, where linear eigenvalue calculations show the strongest planetary wave instability." He suggested that this response was related to the possibility that "the temporal instability of an eigenmode (complex frequency with real wavenumber) may correspond to a spatially unstable radiating mode that is forced by a given frequency (real frequency with a complex wavenumber)." Liu (1999b) as well as Dewar and Huang (2001) carried out similar model perturbation experiments but with a more complex and partially ventilated subtropical gyre flow. Dewar (1998) analyzed the effect of the mean flow on the phase speed of long baroclinic waves in a subtropical gyre. Kubokawa and Nagakura (2002) studied the evolution, in both ventilated and unventilated regions, of disturbances generated by more localized initial disturbances and/or regions of Ekman pumping, both by the Fourier method and numerically. Even in the present idealized case, with no horizontal current shear, Walker and Pedlosky (2002) showed that, with nonzero purely meridional background flow confined to the upper layer of a 2-layer QG model, there was no minimum vertical shear required for instability so that instability need not be confined to any particular region of the subtropical gyre.

c. Restrictions and approximations

Ocean circulation varies over a range of time scales from days to many years. Motivated by a desire to understand the effects of periodic forcing on the underlying and much more slowly varying circulation, the numerical model studies quoted above have typically been carried out by generating a numerical solution retaining all terms including nonlinear ones, first with steady forcing and then additionally imposing the periodic forcing (e.g., Liu 1999b; Dewar and Huang 2001). In contrast, the accompanying supporting and interpretive analytical studies linearize the periodic flow about a background flow that is assumed steady [Liu 1999a,b; Dewar and Huang 2001: the single nonlinear analytical exception known to us is Dewar (1989)]. The solutions of this paper are likewise linearized, about the particularly simple background flow described below.

The overriding motivation for linearization is, of course, the resulting great simplification of analysis. Linearization generally fails when the amplitude of the periodic forcing/flow becomes large, yet often provides insight into the structure of more complete solutions. Thus in the model studies quoted above, the authors show that linearized solutions are very useful in understanding the spatial propagation and evolution of the—in principle nonlinear—perturbations of model solutions that are excited by spatially localized periodic forcing. Note however that it is difficult to assess from these studies the limits of validity of the linearized solutions for the ocean, partly because these model studies do not attempt to closely match the strength of the periodic forcing to that acting on the real ocean.

Investigation of the linearized solutions obtained below was initially motivated by the observation of persistent patterns in wind forcing on scales smaller then the gyre scales (Chelton et al. 2004), and furthermore we focus on the annual period signal. However, it is important to note that the annual period solutions presented here are readily scaled to different periods, as explained in section 3, and the effects on the sympathetic part of the solution of changing the forcing period are discussed in section 4.

Because the annual frequency is one of the most

strongly energized frequencies in the ocean, linearization may be less appropriate than at other frequencies. In defense of linearization at the annual period, it is appropriate to cite the qualitative correspondence between simple linear Rossby wave models driven by annual winds and the sea level signal documented by Chelton and Schlax (1996) and by Wang et al. (2001) in the Indian Ocean. In the final analysis however, particularly for the annual period in the present baroclinically unstable case, linearization must be viewed as a preparation for more complete analysis rather than as definitive.

In this paper, attention is restricted to unventilated background flows. Kubokawa and Nagakura (2002) compare free and forced solutions in the ventilated and unventilated regions of a 2½ layer model subtropical gyre and find that the stability properties of the linearized long-wave solutions are very different between the two regions. This makes it difficult to simply anticipate the results of extension of the present analysis to ventilated regions; such an analysis must be left for a subsequent study.

The most important remaining approximations are (i) the assumption that the dominant scales of the solutions are much smaller than the lateral scales of the subtropical gyre so that the gyre flow may be idealized as spatially constant and (ii) the long-wave approximation. It is important to note that i precludes exploration of the effects associated with lateral variation in the vertical structure of the modes. The long-wave approximation ii restricts the analysis to scales substantially larger than the largest baroclinic Rossby radius of deformation. To simultaneously satisfy both i and ii, the initial scales are chosen to be a good deal smaller (hundreds of kilometers) than the scales characterizing the gyre (thousands of kilometers) yet a good deal larger than the baroclinic Rossby radii (tens of kilometers) and solutions are presented only for times sufficiently small that scales on the order of Rossby radii are not yet strongly energized. While it is difficult to rigorously satisfy these two requirements in a realistic subtropical gyre, it is meaningful to investigate the solution dynamics in detail since, on one hand, the long-wave model is widely used to discuss basic ocean dynamics and, on the other hand, there exist persistent wind features on scales of several hundred kilometers (Chelton et al. 2004).

It has been shown that the fastest-growing waves in the subtropical gyre flow are at the scales of the Rossby radius of deformation (e.g., Spall 1994). This study provides detailed analysis of baroclinic instability acting on larger scales at which the planetary geostrophic approximation is valid. A subsequent and more complete

analysis would include both the gyre scale and the Rossby radius of deformation scales simultaneously. Because of the complexity of this problem, it would have to be addressed numerically. The present detailed analytical study should thus be viewed as necessary preparation for a more complete study that also includes smaller scales.

2. Dynamics

Theories of large-scale ocean flow commonly make use of the planetary geostrophic (PG) approximation in which the open ocean flow is geostrophic except for a surface Ekman layer. A 2½-layer PG model (appendix A) is here supposed to govern the wind-driven steady circulation of a subtropical gyre in a rectangular mid-latitude basin. In this paper we consider small amplitude time-dependent perturbations of such a steady flow that are driven by specified time-dependent surface Ekman pumping $w'_E(x, y, t)$. An important feature of the steady solutions considered here is that, since the deeper layers do not surface to be exposed to the wind, the steady flow is confined to the upper layer. The 2½-layer equations for time-dependent perturbations η_1 and η_2 of the upper and lower interfaces between the three layers of the background flow are (appendix A)

$$\eta_{2t} + U_R \eta_{2x} + V_R \eta_{2y} = \eta_{1t} \quad \text{and}$$

$$\eta_{2t} - c\eta_{2x} + (C\eta_1)_x = w'_E. \tag{1}$$

In these equations and subsequently, subscripts x, y, and t denote differentiation. For notational simplicity we have introduced the abbreviations

$$C(x, y) \equiv \frac{\beta \gamma_1 H_1^0}{f^2} \quad \text{and} \quad c(y) \equiv \frac{\beta \gamma_2 H_{20}}{f^2} \tag{2}$$

in which H_j^0 is the unperturbed depth of the base of layer j (indices $j = 1, 2, 3$ correspond to the upper, middle, and lower layer), the notation $H_{20} \equiv H_2^0$ emphasizes that the depth of the bottom of the unperturbed stagnant middle layer is constant, $f(y)$ is the Coriolis parameter, the γ_j are reduced-gravity parameters defined in appendix A, and

$$U_R \equiv \frac{\gamma_2}{\gamma_1}(u_g + c_R) \quad \text{and} \quad V_R \equiv \frac{\gamma_2}{\gamma_1} v_g, \tag{3}$$

where u_g and v_g are the mean upper layer geostrophic velocity components $-\gamma_1 H_{1y}^0/f$ and $\gamma_1 H_{1x}^0/f$ in the zonal x and meridional y directions and $c_R = C - (\gamma_1/\gamma_2)c$.

3. The initially forced problem

Suppose that both interfacial displacement perturbations η_1 and η_2 are initially zero. Spatially localized Ekman pumping w'_E is turned on at the initial time $t = 0$ and thereafter oscillates at the fixed frequency σ_0. We

call this the initially forced problem. The resulting solution consists of a transient part that satisfies a homogeneous initial value problem and a sympathetic part that oscillates at the frequency of the forcing.

We assume that the gyre scale over which the steady background flow varies horizontally is substantially greater than the horizontal scale of the perturbations. Under this assumption c, C, U_R, and V_R are taken to be constant. For validity of the underlying PG approximation, the horizontal scale of the perturbations must remain larger than the baroclinic Rossby radii characterizing the stratification. This requirement ultimately limits the length of time over which our solutions remain valid approximations.

The governing equation for the time-dependent interface displacement of the middle layer η_2 is then, from (1),

$$CU_R\eta_{2xx} + CV_R\eta_{2xy} + (c + C)\eta_{2xt} - \eta_{2tt} = -(w'_E)_t,$$
$$t > 0 \qquad (4)$$

and η_1 is given in terms of η_2 by either of (1). We ultimately consider a Gaussian patch of Ekman forcing of width L

$$w'_E = E(x, y)e^{-i\sigma_0 t} \quad \text{and} \quad E(x, y) = W_0 e^{-(x^2+y^2)/L^2}.$$
$$(5)$$

The initial conditions are

$$\eta_1(x, y, 0) = 0 \quad \text{and} \quad \eta_2(x, y, 0) = 0. \qquad (6)$$

Note that solutions $\eta_2(x, y, t)$ of (4), (5), and (6) are also solutions $\eta_2(sx, sy, st)$, where s is any real scale factor, provided that σ_0, W_0, and L are scaled by σ_0/s, W_0/s, and sL. The solutions presented subsequently are thus readily scaled up or down in space and time without change of form (providing the scales remain larger than the Rossby radii).

We Fourier transform (4) in space using the conventions

$$\hat{\eta}_2(k, l, t) = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \eta_2(x, y, t)e^{-i(kx+ly)} \, dx \, dy \qquad (7)$$

and

$$\eta_2(x, y, t) = \frac{1}{4\pi^2} \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} \hat{\eta}_2(x, y, t)e^{i(kx+ly)} \, dk \, dl \qquad (8)$$

to obtain

$$[-CU_R k^2 + CV_R kl + (c + C)ik\partial_t - \partial_{tt}]\hat{\eta}_2$$
$$= i\sigma_0 \hat{E}(k, l)e^{-i\sigma_0 t}, \qquad (9)$$

where $\hat{E}(k, l)$ is the Fourier transform of the Gaussian disturbance (5)

$$\hat{E}(k, l) = W_0 L^2 \pi e^{-[L^2(k^2+l^2)]/4}. \qquad (10)$$

The solutions for $\hat{\eta}_1$ and $\hat{\eta}_2$ with Fourier transformed initial conditions (6) consist of a transient (T) part and a sympathetic (S) part that oscillates at the forcing frequency σ_0:

$$\hat{\eta}_1 = \hat{\eta}_{1T} + \hat{\eta}_{1S} \quad \text{and} \quad \hat{\eta}_2 = \hat{\eta}_{2T} + \hat{\eta}_{2S}. \qquad (11)$$

The sympathetic solution is given by the Fourier transform of the first of (1) with harmonic time dependence and by (9) as, respectively,

$$\hat{\eta}_{1S} = \Sigma_0 \hat{\eta}_{2S} \quad \text{and} \quad \hat{\eta}_{2S} = \frac{i\sigma_0 \hat{E}(k, l)e^{-i\sigma_0 t}}{(\sigma_0 - \sigma^+)(\sigma_0 - \sigma^-)},$$
$$(12)$$

where $\sigma^+(k, l)$ and $\sigma^-(k, l)$ are the two solutions of the dispersion relation for plane wave solutions of the homogeneous version of (9),

$$\sigma^2 + \sigma(c + C)k - CU_R k^2 - CV_R kl = 0, \qquad (13)$$

and, from (1), Σ_0 is

$$\Sigma_0 = (\sigma_0 - kU_R - lV_R)/\sigma_0. \qquad (14)$$

The properties of free solutions with dispersion relation (13) have already been discussed by Liu (1999a,b) and by Kubokawa and Nagakura (2002). The two roots of (13) correspond to the fast and slow baroclinic mode, or the "Non Doppler Shift" (N) and the "Advective" (A) modes of Liu (1999a), which reduce to the first and the second internal baroclinic mode in the case of no background flow. When the solutions become unstable, their vertical profiles are no longer necessarily orthogonal, but the decomposition into fast and slow modes implicit in the present method of solution remains unique. The wavenumber dependence of the fast root is summarized in Fig. 1. The properties of the solutions necessary for understanding the evolution of the solutions of the initially forced problem will be invoked as needed in the subsequent text.

The transient solution $\hat{\eta}_{2T}$ satisfies the homogeneous version of (9), so that $\hat{\eta}_{1T}$ and $\hat{\eta}_{2T}$ can be written in the form:

$$\hat{\eta}_{1T}(k, l, t) = \Sigma^+ A e^{-i\sigma^+ t} + \Sigma^- B e^{-i\sigma^- t} \quad \text{and}$$

$$\hat{\eta}_{2T}(k, l, t) = A e^{-i\sigma^+ t} + B e^{-i\sigma^- t} \qquad (15)$$

in which Σ^+ and Σ^- are given, from (1), by

$$\Sigma^+ = (\sigma^+ - kU_R - lV_R)/\sigma^+ \quad \text{and}$$

$$\Sigma^- = (\sigma^- - kU_R - lV_R)/\sigma^-. \qquad (16)$$

The coefficients A and B are obtained from the initial conditions.

The corresponding transient and sympathetic solutions for η_2 in physical space are finally given by
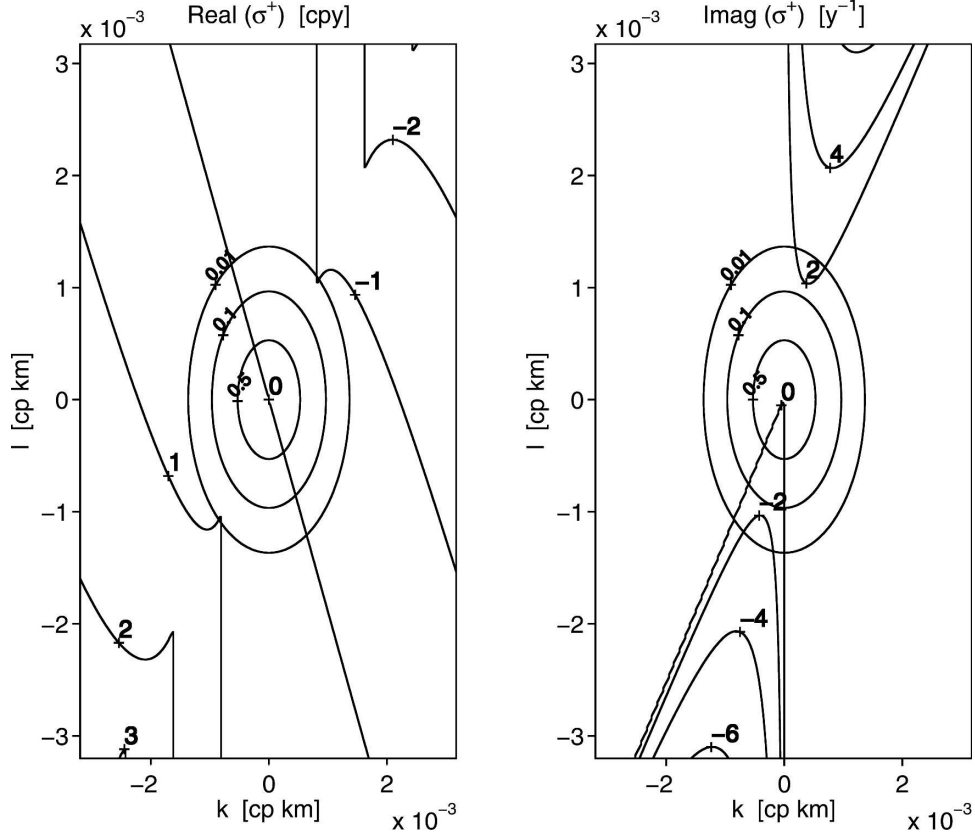
FIG. 1. The dispersion relation (13) for the σ^+ (fast) mode in the presence of the southwestward background flow: contours of (left) $\mathrm{Re}(\sigma^+)$ (yr^{-1}) and (right) $\mathrm{Im}(\sigma^+)$ vs horizontal wavenumber components k and l (cycles km^{-1}). Elliptical lines show contours in wavenumber space of the amplitude of the Gaussian forcing (physical space width 500 km) relative to the amplitude at $k = 0, l = 0$. See text for parameters.

$$\eta_{2T}(x, y, t) = \mathrm{Re}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\hat{E}(k, l)}{(\sigma^+ - \sigma^-)} e^{i(kx+ly)}\left(\frac{-i\sigma^+}{\sigma_0 - \sigma^+} e^{-i\sigma^- t} + \frac{i\sigma^-}{\sigma_0 - \sigma^-} e^{-i\sigma^+ t}\right)\frac{dk}{2\pi}\frac{dl}{2\pi}\right] \tag{17}$$

and

$$\eta_{2S}(x, y, t) = \mathrm{Re}\left[e^{-i\sigma_0 t}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \hat{E}(k, l)e^{i(kx+ly)}\frac{i\sigma_0}{(\sigma_0 - \sigma^+)(\sigma_0 - \sigma^-)}\frac{dk}{2\pi}\frac{dl}{2\pi}\right]. \tag{18}$$

Similar expressions for η_{1T} and η_{1S} are readily obtained from these:

$$\eta_{1T}(x, y, t) = \mathrm{Re}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\hat{E}(k, l)}{(\sigma^+ - \sigma^-)} e^{i(kx+ly)}\left(\frac{-i\sigma^+\Sigma^+}{\sigma_0 - \sigma^+} e^{-i\sigma^- t} + \frac{i\sigma^-\Sigma^-}{\sigma_0 - \sigma^-} e^{-i\sigma^+ t}\right)\frac{dk}{2\pi}\frac{dl}{2\pi}\right] \tag{19}$$

and

$$\eta_{1S}(x, y, t) = \mathrm{Re}\left[e^{-i\sigma_0 t}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \hat{E}(k, l)e^{i(kx+ly)}\frac{i\sigma_0\Sigma^0}{(\sigma_0 - \sigma^+)(\sigma_0 - \sigma^-)}\frac{dk}{2\pi}\frac{dl}{2\pi}\right]. \tag{20}$$

4. Solution of the initially forced problem

a. Numerical solution for transient and sympathetic components

This section discusses solutions $\eta_1(x, y, t)$ and $\eta_2(x, y, t)$ of (1) initiated from rest by Ekman pumping of the Gaussian form (5) initiated at time zero and subsequently periodic with annual frequency σ_0. These have been evaluated both by summation of Fourier components in a periodic domain whose lateral extent is much greater than that of the forcing and by time stepping the finite difference versions of (1) forward in the same domain. In the Fourier method, the initial conditions and forcing are decomposed into Fourier components $e^{i(kx+ly)}$, each Fourier component is independently advanced in time according to its dispersion relation $\sigma = \sigma(k, l)$ by multiplication by $e^{-i\sigma(k,l)t}$, and the Fourier components are then recombined as in (17) through (20) to give the solutions at time t.

The background flow is that prevailing in the eastern part of a steady and unventilated subtropical gyre occupying a midlatitude basin of meridional extent 3000 km and zonal extent 10 000 km, driven by steady Ekman pumping of the form (52). The resting depths (before application of the steady Ekman pumping) of the base of two upper layers of the model H_{10} and H_{20} are chosen to be 500 and 1000 m with reduced-gravity parameters (appendix A) $\gamma_1 = 0.0245$ and $\gamma_2 = 0.01$ m s^{-2}. At the location at which the time-dependent perturbation wind patch (5) is centered (1000 km west of the eastern boundary of the basin and 1000 km north of its southern boundary), the corresponding wave scale speeds C and c [(2)] are, respectively $C = 4.60$ and $c = 3.23$ cm s^{-1}. For the particular example shown here, the coefficients U_R and V_R appearing in (1) have the values $U_R = -2.3$ cm s^{-1} and $V_R = -1.9$ cm s^{-1}. The zonal component of the geostrophic velocity at this location is $u_g = -2.8$ cm s^{-1} and the meridional component of the geostrophic velocity $v_g = -4.7$ cm s^{-1}, so the angle between the mean flow and the zonal direction is 239°. Figure 1 shows that the direction of the wave vector k, l associated with the fastest temporal growth $\text{Im}(\sigma^+(k, l))$ for the root $\sigma^+(\sigma^-)$ of (13) is 59° (239°) so that the mean flow direction determines the direction along which the fastest-growing Fourier component wavenumbers lie. The direction of the wavenumber vector of the fastest-growing modes is determined by the competition between maximal energy release for a wave vector oriented parallel to the mean flow and the β effect, which stabilizes the flow for the perturbations across the mean potential vorticity gradient (Pedlosky 1987). The fact that the direction of the wavenumber vector of

the fastest-growing mode is almost parallel to the direction of the mean flow suggest that the vertical shear is strong enough to overcome the stabilizing effect of β. The Rossby radii are 41 and 17 km, values comparable to those in the related work of de Szoeke and Chelton [1999, (2.11)].

Figure 2 shows the Fourier summation solution for η_1 and η_2 at 500, 900, and 1200 days after the initiation of forcing. The solutions after 2000 days are shown in the upper panels of Fig. 4. The origin of coordinates $x = 0$, $y = 0$ in the figure is the center of the region of forcing. The background flow and forcing parameters are those of Fig. 1. After 900 days a well defined beam of stable waves is directed slightly toward the north of west away from the region of forcing, but the full Fourier solution contains an additional transient feature (17) and (19) that travels westward and clearly grows substantially from 900 days to 2000 days. Careful examination of Fig. 2 shows that the in-phase nature of the fluctuations of η_1 and η_2 that prevails in the beam of the sympathetic solution is replaced by an order 90° phase shift in the growing region. The growth is clearly the result of baroclinic instability of the background steady flow, already noted by Liu (1999a), Dewar and Huang (2001), and Kubokawa and Nagakura (2002).

Numerical evaluations (not shown) of the Fourier solution for the beam indicate that, as the frequency of Ekman pumping is increased, the distance between extrema in the beam decreases and the amplitude of the beam also decreases, but the direction of the beam is not visibly frequency dependent. For the present parameters, radiation in the beam is virtually extinguished for forcing frequencies of 0.5 cpy or higher. It is clear from the dispersion relation (13) that increasing the frequency decreases the wavelength of the sympathetic solution, but simple arguments do not predict the direction of the beam nor do they explain the decrease in beam intensity at high frequencies.

The Fourier summation and finite difference methods are very flexible but, on account of necessarily limited spatial resolution and "wrap around" in the periodic spatial domain, they are not well suited to answer a number of questions raised by the solutions of Fig. 2. Those questions are:

(i) Is the beam appearing in Fig. 2 the entire sympathetic solution?
(ii) Why is the beam so strongly directional?
(iii) Why is radiation at the forcing frequency strongly inhibited at higher frequencies?
(iv) What is the long-time structure of the transient solution?
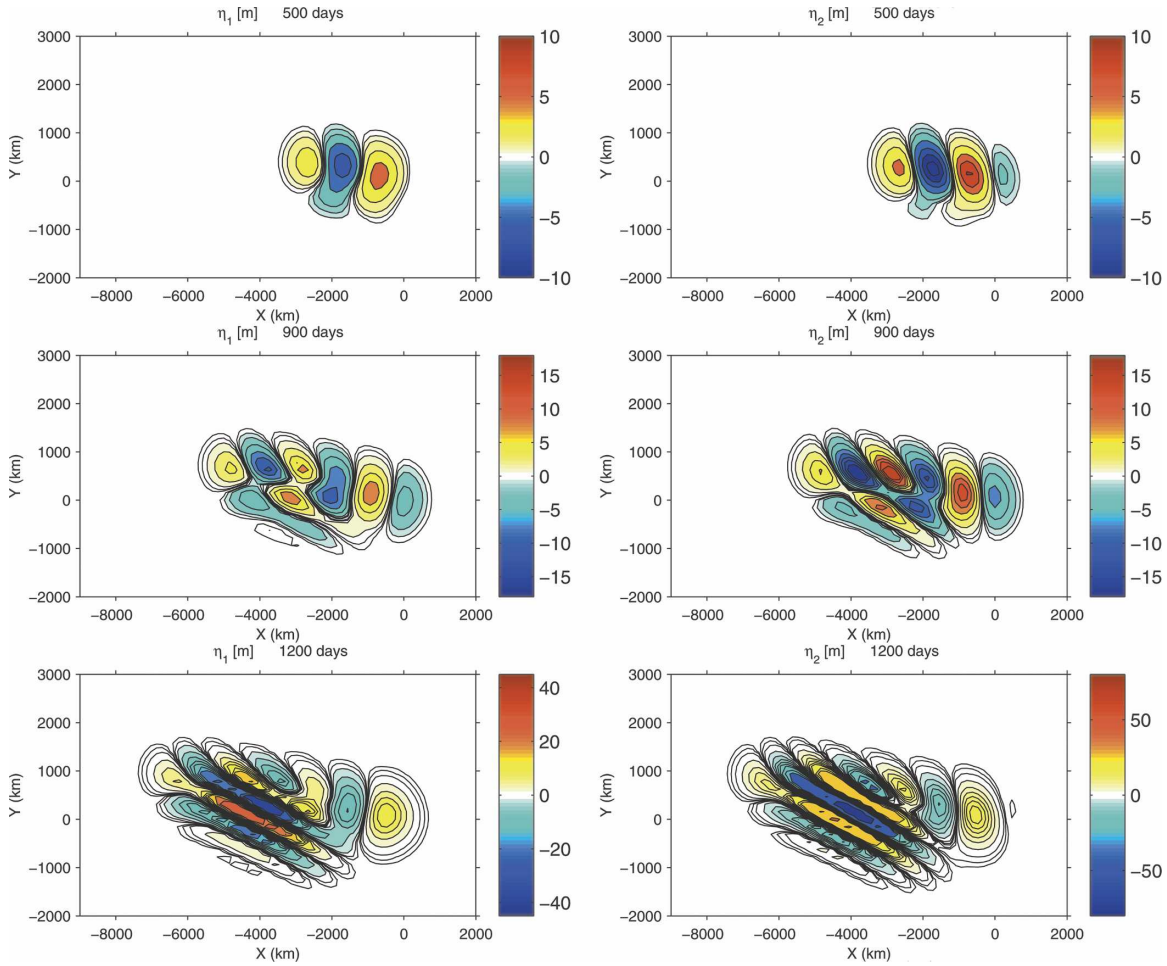(v) Is the instability of the transient solution convec-

FIG. 2. A snapshot of the upper- and the middle-layer Fourier summation interface displacement (η_1, η_2: m) for the total solution (sympathetic plus transient) at times 500, 900, and 1200 days after initiation of forcing. The parameters describing the background flow are those of Fig. 1. Contour levels are 80, 50, 25, 15, 10, 5, 2, 1, 0.5, and 0.25 m.

tive in the manner defined by Huerre and Monkewitz (1990)?

To answer these questions it is necessary to develop solutions free of the shortcomings of the Fourier method. We therefore next work out analytical approximations to the solutions valid for long times after the initiation of forcing.

b. The sympathetic solution

The foregoing Fourier component summation solution of the initially forced problem consists of a transient, (17) and (19), and a sympathetic, (18) and (20), part. It was seen in Fig. 2 that this transient solution may become unstable. The fact that the plane wave dispersion relation (13) may have complex wavenumbers k and l for real frequencies σ suggests, as noted by Liu (1999a), that the sympathetic solution might grow spatially. Does it do so?

The full sympathetic solution, with allowance for possible spatial growth, may be found by a highly simplified version of the method of Briggs (1964, chapter 2: 8–46). If we rewrite the Fourier transform (10) of the Gaussian disturbance (5) as $\hat{E}(k, l) = \hat{E}(k)\hat{E}(l)$ with $\hat{E}(k) = L\sqrt{\pi W_0}e^{-k^2L^2/4}$, and so on, then the sympathetic solution (18) may be rewritten as

$$\eta_{2S}(x, y, t) = \mathrm{Re}\left[\int_{-\infty}^{\infty} \hat{E}(l)e^{ily}\eta_{2S}^x(x, l)\frac{dl}{2\pi}\right], \quad (21)$$

where

$$\eta_{2S}^x(x, l) \equiv \left[\int_{-\infty}^{\infty} \frac{i\sigma_0}{(\sigma_0 - \sigma^+)(\sigma_0 - \sigma^-)}\hat{E}(k)e^{i(kx-\sigma_0 t)}\frac{dk}{2\pi}\right].$$
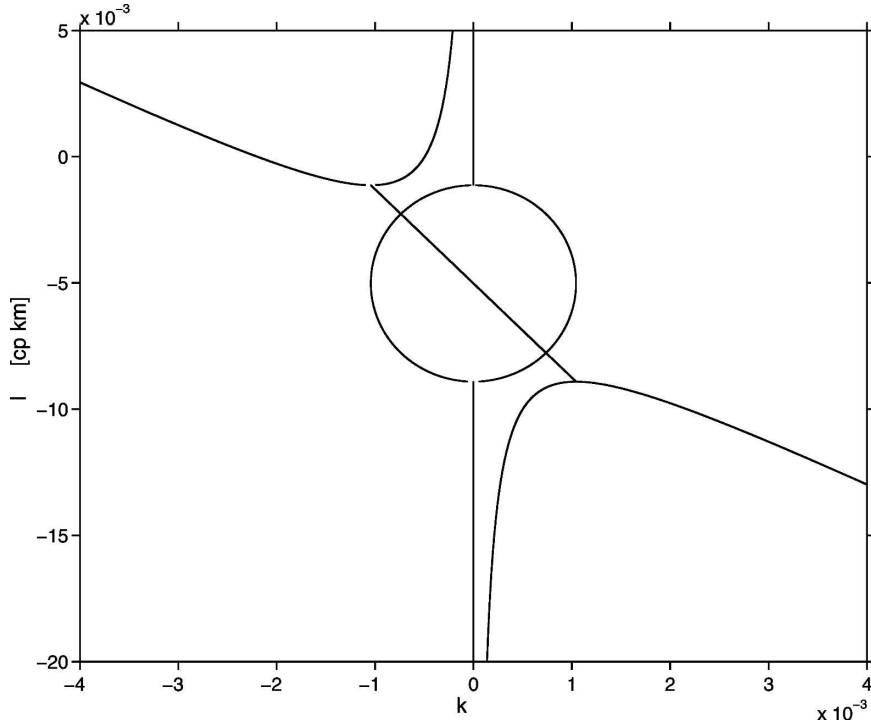
$$(22)$$

FIG. 3. For the parameters of Fig. 1, the dispersion relation (13) is solved for zonal wavenumber k assuming that the meridional wavenumber l and the frequency σ_0 are real, and the results are plotted in the k, l plane (km^{-1}). The hyperbolic lines show the real part of k, and the vertical line with circle shows the imaginary part of k. For $l_{lo} < l < l_{up}$, k is complex. As noted by Liu (1999a), complex values of k may correspond to spatial growth.

Here $\eta^x_{2S}(x, l)$ may be written in terms of a Green's function G_2 as

$$\eta^x_{2S}(x, l) = \int_{-\infty}^{\infty} G_2(x - x')e^{-x'^2/L^2}\, dx' \qquad (23)$$

with similar expressions for $\eta^x_{1S}(x, l)$ and G_1. From (1), G_1 and G_2 obey the coupled equations

$$-i\sigma_0 G_2 + U_R G_{2x} + ilV_R G_2 = -i\sigma_0 G_1 \quad \text{and}$$
$$- i\sigma_0 G_2 + cG_{2x} - CG_{1x} = \delta(x). \qquad (24)$$

Here G_2 is continuous across the origin $x = 0$ but G_1 jumps discontinuously in the amount $-1/C$ as x increases across the origin: G_1 and G_2 are of the form $Ae^{ik^+ x} + Be^{ik^- x}$ away from the origin, where A and B are constants and k^+ and k^- are the solutions of the quadratic equation in (13) with real l and with frequency σ_0 (the superscript signs correspond to the sign choice in the quadratic formula in solving for k). Once A and B have been chosen, then the integrations in (21) and (23) may be done numerically.

The solutions k^+ and k^- are plotted in Fig. 3 for real l and for the annual frequency σ_0. They are complex

when l lies between two critical meridional wavenumbers l_{up} and l_{lo} (where $l_{up} > l_{lo}$). We thus consider the three separate cases $l > l_{up}$, $l < l_{lo}$, and $l_{up} > l > l_{lo}$. For $l > l_{up}$ both k^+ and k^- are real, and may be shown to correspond to free waves with westward group velocity. The radiation condition thus requires that

$$G_2(x) = Ae^{ik^+ x} + Be^{ik^- x},$$
$$G_1(x) = \Sigma^+ Ae^{ik^+ x} + \Sigma^- Be^{ik^- x}, \quad x < 0, \quad \text{and}$$
$$G_2(x) = G_1(x) = 0, \quad x > 0. \qquad (25)$$

The constants A and B are fixed by the continuity of G_2 and the jump of G_1 across the origin noted above. For $l < l_{lo}$ both k^+ and k^- are again real and may be shown to correspond to free waves with eastward group velocity for k^+ and westward group velocity for k^-. The radiation condition thus requires that

$$G_2(x) = Ae^{ik^- x}, \quad G_1(x) = \Sigma^- Ae^{ik^- x}, \quad x < 0,$$

and

$$G_2(x) = Be^{ik^+ x}, \quad G_1(x) = \Sigma^+ Be^{ik^+ x}, \quad x > 0. \qquad (26)$$

Again the constants A and B are fixed by the continuity of G_2 and the jump of G_1 across the origin noted above.

For the particular parameters of the solutions of this paper, very little energy appears in meridional wavenumbers $l < l_{lo}$ on account of the small amplitude of the Gaussian forcing at meridional wavenumbers $l < l_{lo}$ in wavenumber space.

For $l_{lo} < l < l_{up}$ both k^+ and k^- are complex, and the radiation condition no longer determines the form of the solution on either side of the origin. Pedlosky (2005, personal communication) has pointed out that, in this case, a heuristic version of the procedure of Briggs (1964) may be used to obtain the form of the solution away from the origin. The procedure is to add artificially large damping ($-i\sigma_0 \rightarrow -i\sigma_0 + \varepsilon$, ε large) and to determine the form of the solution (the constants A and B) on either side of the origin by retaining only those terms that correspond to dissipatively dominated decay away from the origin. Once this has been done, let the damping decrease to zero without changing A or B. If a solution that decays away from the origin when dominated by large dissipation grows in the same direction when the dissipation is sufficiently reduced toward zero, then that solution is spatially unstable and must be retained. This is the frictional equivalent of the causality condition enforced by requiring the radiation condition in the case where k^+ and k^- were real. In the present case, when the substitution, $-i\sigma_0 \rightarrow -i\sigma_0 + \varepsilon$, is made in the dispersion relation (13), then for large ε both k^+ and k^- correspond to westward decay but as $\varepsilon \rightarrow 0$, k^+ switches to westward growth. Consequently, for $l_{lo} < l < l_{up}$,

$$G_2(x) = Ae^{ik^+x} + Be^{ik^-x},$$

$$G_1(x) = \Sigma^+ Ae^{ik^+x} + \Sigma^- Be^{ik^-x}, \quad x < 0, \quad \text{and}$$

$$G_2(x) = G_1(x) = 0, \quad x > 0, \tag{27}$$

and the constants A and B are fixed by the continuity of G_2 and the jump of G_1 across the origin noted above. The result is that the sympathetic solution is spatially unstable and grows westward.

Because meridional wavenumbers for which $l < l_{lo}$ make effectively no contribution to the present solution, that solution may also be obtained by making a finite difference approximation to the y derivative in (1) and then integrating the resulting equations with harmonic time dependance westward from the eastern boundary. Parameters such as domain size, etc. enter the finite difference code in a simpler manner than the code that evaluates (21)–(27) so that the solutions for $\eta_1(x, y)$ and $\eta_2(x, y)$ shown in the lower panels of Fig. 4 for the parameters and domain of Fig. 2 were actually obtained by the finite difference method.

Figures 2 and 4 make clear that, as the full solution evolves, ever higher amplitudes are found ever farther west and south of the initially forced region. Comparison of the full solution (Fig. 4, upper panels) after 2000 days with the sympathetic solution (Fig. 4, lower panels) shows that, as the most rapidly growing region moves westward, it "leaves behind" the sympathetic solution, first the narrow stable beam (composed of Fourier components with meridional wavenumbers $l > l_{up}$) and then the westward growing part (composed of Fourier components with meridional wavenumbers $l_{lo} < l < l_{up}$). The latter is just beginning to be clearly visible by day 2000.

The sympathetic solution $\eta_2(x, y)$ given by (21)–(27) may be thought of as a Green's function convolved in x and y with the Gaussian forcing. In the course of solving (21)–(27) when $l > l_{up}$, that Green's function (not shown) could be constructed by letting the factors $\hat{E}(l)$ and $\hat{E}(k)$ in (21) and (22) be constant; it was a complicated function of position characterized by very different local wavelengths at different locations. When convolved in x and y with the Gaussian forcing, this complicated pattern became the simple beam pattern of Fig. 4 because the very anisotropic dispersion relation (13) forces the longest waves to be found in only one particular direction, that of the beam, and it is only those largest-scale waves that survive the smoothing associated with the convolution. As the frequency is increased, the dispersion relation (13) indicates that the waves become shorter and are consequently more strongly attenuated by the convolution; that is why increasing the frequency ultimately extinguishes the beam.

c. Stationary phase approximation to the transient solution

To be able to answer the questions posed when analyzing the numerical solution at the end of section 1, in particular the question of whether the instability observed in the solution is absolute or convective, we next construct an analytical approximation to the transient solution.

The initial conditions satisfied by the transient solution (17) must cancel the sympathetic solution (18) evaluated at the initial time. But, the sympathetic solution evaluated at the initial time is a complicated and spatially extended function of x and y. To more clearly understand the evolution of the transient solution, we instead consider more compact initial conditions

$$\eta_2(x, y, 0) = \eta_{20}(x, y) \quad \text{and} \quad \eta_1(x, y, 0) = \eta_{10}(x, y) \tag{28}$$

in which both $\eta_{10}(x, y)$ and $\eta_{20}(x, y)$ are taken proportional to the Gaussian $e^{-(x^2+y^2)/L^2}$ of (5). The solution
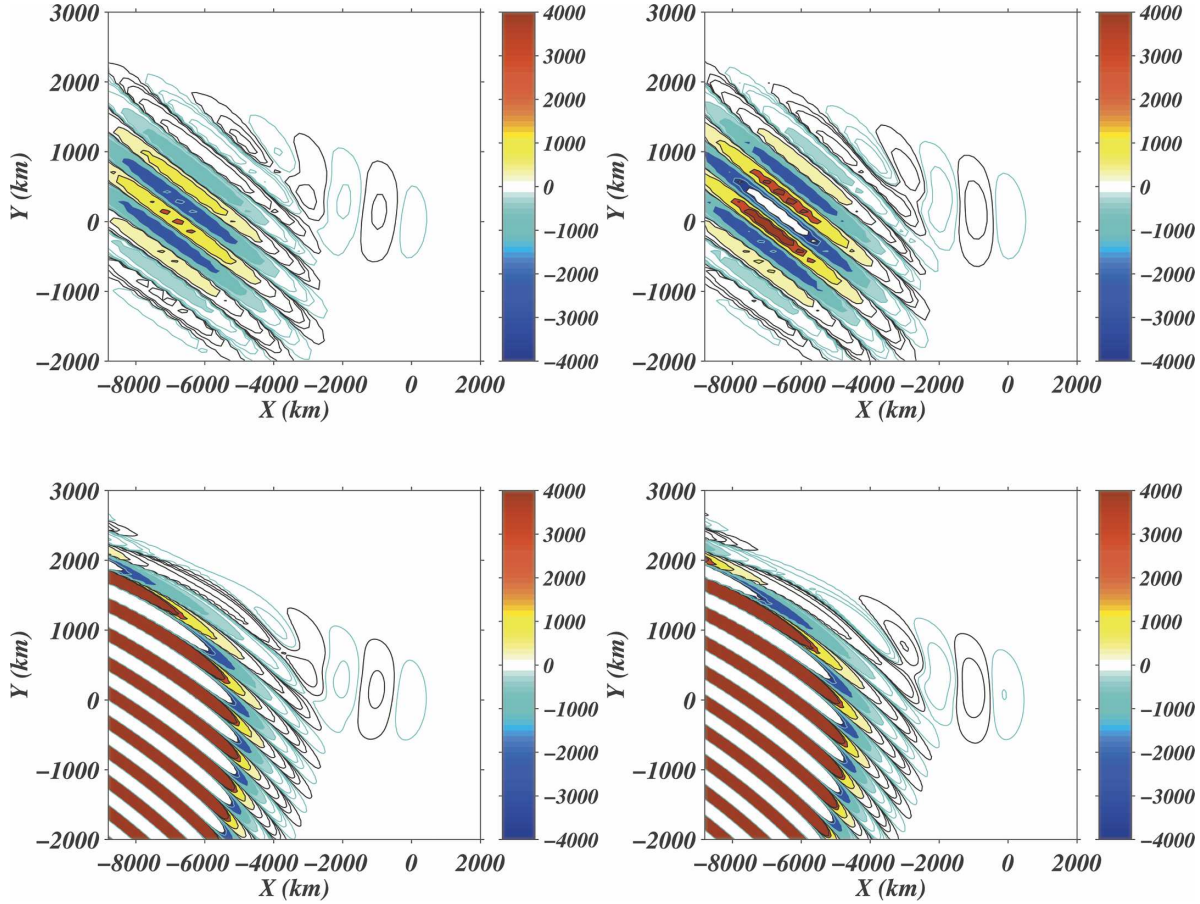
FIG. 4. (top) Snapshot of the upper- and middle-layer Fourier summation interface displacement (η_1, η_2: m) for the total solution (sympathetic plus transient) at time 2000 days after initiation of forcing. (bottom) As in top but for the sympathetic solution. The parameters describing the background flow are those of Fig. 1. Contour levels are 4000, 3000, 1000, 300, 100, 30, 10, 5, and 1.0 m; amplitudes at the far western edge of the bottom panels are on the order of 10^{10} m.

for $\hat{\eta}_{2T}(k, l, t)$ is still of the form (15) in which the coefficients A and B are given by

$$\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} \Sigma^+ & \Sigma^- \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\eta}_{10} \\ \hat{\eta}_{20} \end{pmatrix}. \qquad (29)$$

Inverse Fourier transformation of (15) then furnishes η_{1T} and η_{2T}.

We solve the homogeneous initial value problem asymptotically for large time using the method of steepest descents. The classical asymptotic inverse Fourier transformation (8) for the initial value problem as $t \to \infty$ is given by a sum of terms each one of the form

$$\eta_2 \propto \frac{\exp\left[i\left(k_0\dfrac{x}{t} + l_0\dfrac{y}{t} - \sigma_0\right)t\right]}{t\sqrt{\det\left(\dfrac{\partial^2\Omega}{\partial k_i \partial k_j}\right)}} G(k_0, l_0) \qquad (30)$$

where $\sigma_0 = \Omega(k_0, l_0)$ denotes the dispersion relation (13), $\nabla_k\Omega_0 = (x/t, y/t)$ defines the stationary phase points (e.g., Carrier et al. 1966), and $G(k, l)$ gathers together all the remaining k and l dependence of the integrand. It can easily be shown from (13) that

$$\det\left(\frac{\partial^2\Omega}{\partial k_i \partial k_j}\right) = 0$$

everywhere in wavenumber space and the form (30) fails in general. (This comes about because the long wave assumption has been made.) Nevertheless, there is an orientation of coordinates that makes the matrix

$$\left(\frac{\partial^2\Omega}{\partial k_i \partial k_j}\right)$$

diagonal and one of the eigenvalues is nonzero; we thus may integrate in the direction of the nonzero eigen-

value. We accordingly work in cylindrical polar coordinates; the wavenumber vector components k and l are then described in terms of the magnitude of the wavenumber vector κ and the angle α that the wavenumber vector makes with the zonal direction. We first perform the integration over the angle α by the method of steepest descents. Thereafter, for the present special case of a Gaussian initial disturbance we may perform the integration over the wavenumber magnitude κ again by the method of steepest descents. The radial profile of the cylindrically symmetric initial disturbance thus influences only the result of the last integration.

The middle layer interface displacement $\eta_{2T}(x, y, t)$ is then a sum of terms of the form

$$\left[L^2\pi \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(k^2 + l^2)L^2/4} e^{i(kx+ly-\sigma t)} G(k, l) \frac{dk}{2\pi} \frac{dl}{2\pi} \right]$$

(31)

in which σ is either σ^+ or σ^-. In polar coordinates κ, α this term becomes

$$\text{Re}\left[L^2\pi \int_0^{\infty} \int e^{-\kappa^2 L^2/4} e^{i\kappa Ct\Phi(\alpha)} G(\kappa, \alpha) \frac{\kappa \, d\kappa \, d\alpha}{2\pi \, 2\pi} \right], \quad (32)$$

where Φ is the phase given by

$$\Phi = \xi\cos(\alpha) + \zeta\sin(\alpha) - H\cos(\alpha) \quad (33)$$

and we have introduced

$$\xi = \frac{x}{Ct} + \frac{a}{2}, \quad \zeta = \frac{y}{Ct}, \quad \text{and} \quad H = \frac{\sigma}{C\kappa\cos\alpha} + \frac{a}{2} \quad (34)$$

with $a = 1 + c/C$. Here ξ and ζ are scaled horizontal coordinates that move westward with the frontal feature of the disturbance (defined below).

The α integration in (32) may be carried out using the method of steepest descents for large time t. The stationary phase points $\alpha_0(\xi, \zeta)$ are solutions of

$$\Phi_\alpha = (-\xi\vartheta + \zeta + H\vartheta - H_\alpha)\cos(\alpha) = 0 \quad (35)$$

in which $\vartheta \equiv \tan(\alpha)$. To solve this transcendental equation for $\alpha_0(\xi, \zeta)$, make use of the dispersion relation (13) written in terms of H,

$$H^2 = b + v\vartheta \quad (36)$$

in which

$$b \equiv \frac{a^2}{4} + u, \quad u \equiv \frac{U_R}{C}, \quad \text{and} \quad v \equiv \frac{V_R}{C}. \quad (37)$$

Eliminating reference to ϑ and H between (35) and (36) yields a quartic whose coefficients have no explicit dependence on α,

$$H^4 - 2H^3\xi + 2H(b\xi + v\zeta) - (b^2 + v^2) = 0. \quad (38)$$

Solutions of this are the quantity

$$H_0 = \frac{\sigma}{C\kappa\cos(\alpha_0)} + \frac{a}{2}$$

evaluated at the stationary phase points $\alpha = \alpha_0$. The value of α_0 itself then readily follows from (36). The stationary phase values α_0 may be complex, corresponding to growth or decay of their contribution to the solution. The term (32) thus becomes

$$\text{Re}\left[L^2\pi \int_0^{\infty} e^{-\kappa^2 L^2/4} e^{i\kappa Ct\Phi_0} e^{i\varepsilon(\alpha_0)} \frac{G(\kappa, \alpha)}{\sqrt{|\kappa Ct\Phi_{0\alpha\alpha}|}} \frac{\kappa \, d\kappa}{2\pi} \right], \quad (39)$$

where Φ_0 denotes $\Phi(\alpha_0)$, $\varepsilon(\alpha_0) = \pi/4 - \arg(\Phi_{0\alpha\alpha})/2$, and the phase of $\Phi_{0\alpha\alpha}$ is chosen so that $-\pi/2 < \Phi_{0\alpha\alpha} < 3\pi/2$. The κ integration is now readily carried out by stationary phase when t is large to finally obtain

$$\eta_{2T}(x, y, t)$$

$$\approx \text{Re} \sum_{\alpha_0} \left[\frac{L\sqrt{2}}{4} e^{i\varepsilon(\alpha_0)} \frac{e^{-\frac{C^2t^2\Phi_0^2}{L^2}} \kappa_0 G(\kappa_0, \alpha_0)}{\sqrt{|\kappa_0 Ct\Phi_{0\alpha\alpha}|}} \right] \cdots$$

(40)

in which κ_0 is the stationary phase point

$$\kappa_0 = \frac{2itC\Phi_0}{L^2} \quad (41)$$

and in which the sum reminds us that there are generally several stationary phase points for each field point ξ, ζ.

1) FRONT

Expression (40) shows that the maximum value of the solution, initially the peak of the initial Gaussian, remains very near the curve in the ξ, ζ plane defined by

$$\Phi_0 = \xi\cos(\alpha_0) + \zeta\sin(\alpha_0) - H_0\cos(\alpha_0) = 0, \quad (42)$$

for times sufficiently short that the solution is not dominated by growth associated with complex stationary phase values α_0. We call this curve the front (Fig. 5). To obtain an equation for the front whose coefficients are not functions of $\alpha_0(\xi, \zeta)$, use (36) to eliminate $H_\alpha(\alpha_0)$
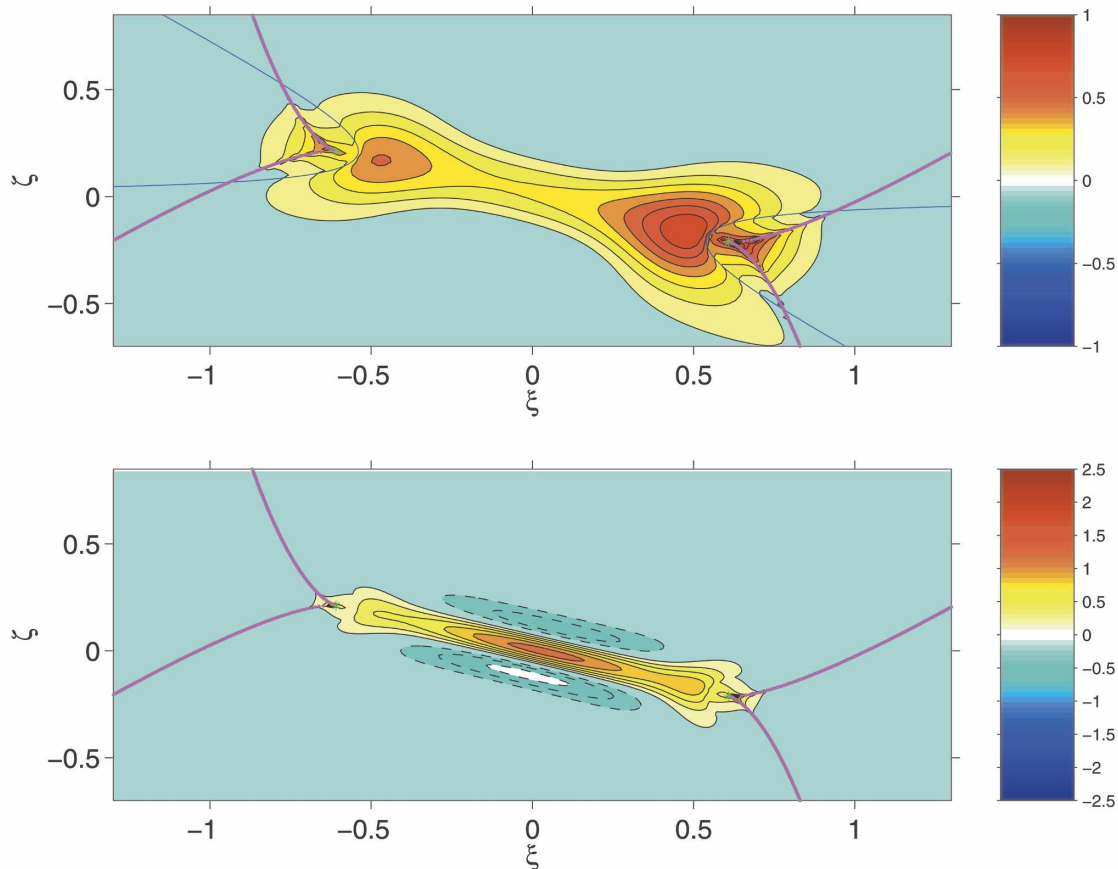
FIG. 5. (top) Snapshot of the time-dependent interface displacement of the middle layer (η_{2T}; m) for the transient solution obtained by stationary phase after 500 days for the Gaussian initial conditions $\eta_{20}(x, y) = e^{-(x^2+y^2)/L^2}$ and $\eta_{10}(x, y) = 0$ with $L = 500$ km. The largest amplitude occurs approximately at the zonal position where the slow mode contribution to the solution would have been centered without background flow, with a strong secondary maximum further to the west at the zonal position where the fast mode contribution to the solution would have been centered without background flow. The maximum amplitude of η_2 is about 1.1 m, whereas the maximum initial amplitude of η_2 was 1.0 m. The light solid line is the front (43). Contour intervals are from 0.1 to 1.0 m with an increment of 0.1 m. (bottom) Same as top but for 900 days. The light solid line is the caustic (46); the solution grows in the region between the caustics. In the middle of the group, the gradient of elevation is nearly parallel to the central axis of the wedge of unstable wavenumbers of Fig. 1. The maximum amplitude of η_2 is about 1.8 m; contour intervals are from -0.6 to 1.8 m with an increment of 0.2 m. In both panels solid contours are positive; dashed contours are negative.

from (35) and combine with (42) in the form $\xi + \zeta\vartheta_0 - H_0 = 0$ to find $\zeta = v/(2H_0)$. Now use this and (36) to eliminate reference to ϑ_0 and to H_0 from (42) to finally obtain the equation for the front as

$$\xi = \frac{v}{4\zeta} + \frac{\zeta b}{v} \qquad (43)$$

in which v is defined in (36). For large $|\xi|$, the front asymptotes to $\xi = \zeta b/v$ and to $\zeta = 0$. It has focii at $\zeta = \pm\sqrt{-v^2 b/4}$, $\xi = 0$. Since the coordinates ξ and ζ move westward at speed $(c + C)/2$, the frontal pattern of Fig. 5 also propagates westward without change at this speed, even though the full solution changes form as it evolves in the moving coordinates.

2) CAUSTICS

The solution (40) displays caustics, shown in Fig. 5. They represent a boundary between a region with a complicated wave pattern where groups of waves interfere and a neighboring region where two fewer groups of waves (possibly none) interfere (Lighthill 1978). At a caustic the first and the second derivatives of the phase vanish:

$$\Phi'(\alpha) = 0 \quad \text{and} \quad \Phi''(\alpha) = 0, \qquad (44)$$

causing the stationary phase solutions to break down at the caustics. In principle this local difficulty may be overcome by using the Airy integral to produce a

"healed" version (Lighthill 1978). That has, however, not been done for the solutions discussed below (Fig. 5).

The first of the equations in (44) may be solved to find the stationary phase values α_0 of α as in the previous section yielding (35). For each stationary phase value of α_0, (44) constitute two equations for ξ and ζ whose solutions define the caustic curves. Insert ξ obtained from the first of the expressions (44) into the second one to obtain ζ as

$$\zeta = \frac{3v^2\vartheta_0 + v^2\vartheta_0{}^3 + 2vb}{4H_0^3}, \tag{45}$$

where H_0 is defined by the dispersion relation in the form (36) with $\alpha = \alpha_0$. From (45) and (44) we get

$$\xi = \frac{v^2(1 + \vartheta_0^2)^2}{4H_0^3} + H_0 - \zeta\vartheta_0. \tag{46}$$

The caustics, asymptote at large distances ξ, ζ to $\pi/2$ and $\arctan(-b/v)$. The caustics meet in cusps at points of the ξ, ζ plane where the first three derivatives of phase with respect to the wavenumber vector direction α are zero:

$$\Phi'(\alpha) = 0, \quad \Phi''(\alpha) = 0, \quad \text{and} \quad \Phi'''(\alpha) = 0. \tag{47}$$

There are two such points, given by $\vartheta = -b \pm (b^2 + v^2)^{1/2}/v$; ξ and η are given by (46) and (45).

Between the caustic curves of Fig. 5 the parts of the solution corresponding to the complex stationary phase points grow in time for the present baroclinically unstable case. In the remainder of the domain however, outside of the caustic curves, all the stationary phase points the contribute to the solution are real and do not result in growth. In this manner, the confinement of complex stationary phase points to just a part of the ξ, η plane causes the growing part of the solution to be restricted to just a part of the x, y plane at any time so that the instability is convective in the manner defined by Huerre and Monkewitz (1990) rather than absolute.

3) DISCUSSION OF STATIONARY PHASE SOLUTION

The initial conditions are of the form (28) with the particular choice $\eta_{20}(x, y) = e^{-(x^2+y^2)/L^2}$ and $\eta_{10}(x, y) = 0$. This choice insures that both fast and slow modes are strongly energized in the solution. The choice of stationary phase points is discussed in appendix B. As noted above, the Gaussian scale is chosen to be $L = 500$ km, whereas the Rossby radii are 41 and 17 km.

The stationary phase solutions (40) for η_2 are displayed in Fig. 5 in the coordinates ξ, ζ at two different times. A point fixed in those coordinates moves westward with speed $(c + C)/2$ and radially outward at speed C in the physical space coordinates x, y. Conse-

quently much (but not all) of the evolution of the solution in physical space coordinates x, y is just westward drift at speed $(c + C)/2$ and radial expansion at speed C. These stationary phase solutions are in good agreement with their less well resolved Fourier counterparts except in the vicinity of the caustics. We may now answer a number of questions about the transient solution. 1) Why does the dominant part of the transient solution quickly elongate itself north-of-west to south-of-east? 2) Why does the solution evolve from the initial Gaussian to a pattern with two distinct maxima (Fig. 5a) to a pattern with a single central maximum (Fig. 5b)?

In answer to question 1, initially the distribution of energy in wavenumber space is isotropic and so is the solution. As Fourier components whose wavenumbers lie in the unstable wedges of Fig. 1 grow, however, the solution increasingly consists of waves whose crests and troughs are oriented north-of-west to south-of-east at right angles to the axis of the wedge of unstable wavenumbers of Fig. 1 and its counterpart for σ^- (not shown). For question 2, as the solution initially evolves (Fig. 5a) the largest amplitude for η_2 occurs approximately at that zonal position in the basin where the slow mode contribution to the solution would have been centered in the absence of background flow, but there is a strong secondary maximum farther to the west at which the fast mode contribution to the solution would have been centered in the absence of background flow. The same two extrema appear in η_1 (not shown) but with different relative amplitudes. Kubokawa and Nagakura (2002) obtained a similar pattern by summing Fourier modes; whereas they discuss the solution in different parts of the ocean, we focus on a more complete description of the space and time variation of the solution. The maximum amplitude of η_2 for the solution of Fig. 5a is about 1.1m whereas the maximum initial amplitude of η_2 was 1.0 m; it is thus clear that baroclinic instability is not yet an important factor in the evolution of the solution.

At some time, when solving by summing Fourier modes, the smallest-scale resolved components are energized. At that point they can no longer interfere with yet-smaller-scale components in the same manner as they do in the fully resolved solution so that the shortest resolved components spuriously dominate the numerical solution, giving it a "checkerboard" texture. The stationary phase solution does not have such a short wave resolution limit. Figure 5b thus shows the stationary phase solution at a time appreciably later than that of Fig. 5a. Now the maximum amplitude is significantly greater than that of the initial condition, indicating that baroclinic instability has become a more important factor in the evolution of the solution.
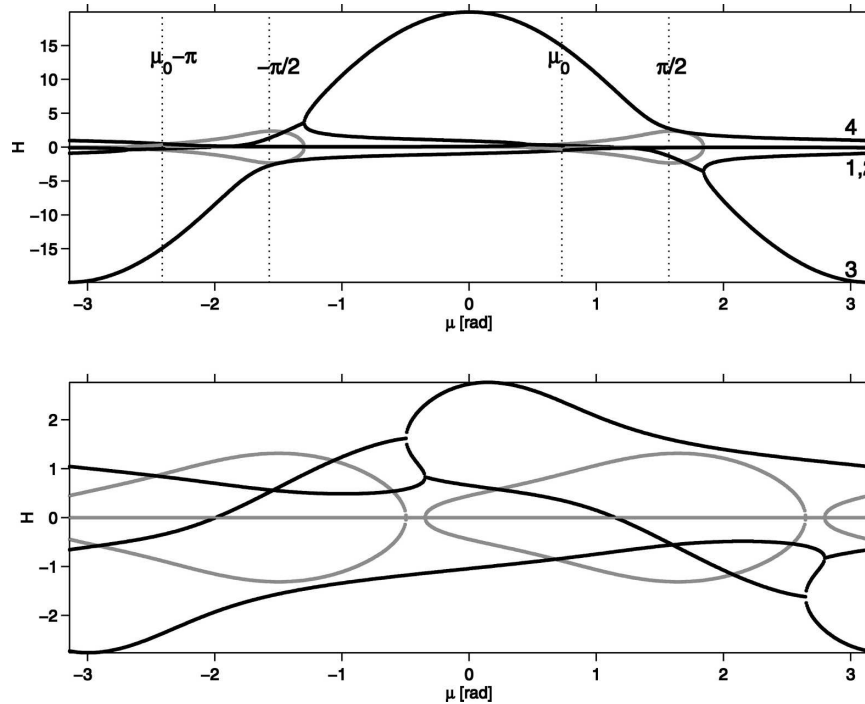
FIG. 6. Four roots H_{0i} of the quartic equation (38) for the quantity $H_0 = [\sigma/C\kappa \cos(\alpha_{0i})] + (a/2)$ in which the α_{0i} are the stationary phase points of the α integration of (32) along two circles of radius R [(top) large $R = 10$ and (bottom) small $R = 1.5$] in the ξ, ζ plane as a function of angle $\mu = \tan^{-1}(\zeta/\xi)$. The real and imaginary part of H_0 are shown as black and gray solid lines, respectively.

By whatever method they have been obtained, the solutions remain valid providing that the dominant horizontal scale are larger than the baroclinic Rossby radii. The limiting time scale at which this ceases to be true has not yet been reached in the solutions of Fig. 5.

What is the nature of the different wave groups that interfere to produce the full transient solution? The stationary phase angles $\alpha_0(\xi, \zeta)$ may be deduced from the quantity H defined in (35). Figure 6 shows H obtained by solving the quartic (38) for ξ, ζ lying on two circles (of radius $R = 10$ and $R = 1.5$) centered at the origin of coordinates in the ξ, ζ plane as a function of the angle $\mu = \tan^{-1}(\zeta/\xi)$. The four angles associated with the two caustics shown in Fig. 5 are clearly visible. Also shown are large R approximations to the locations of the caustics (appendix C).

At large R (Fig. 6a), there are four real stationary phase points $\alpha(\xi, \zeta)$ for the largest values of μ. As μ decreases through successive caustic angles, one pair of these coalesces into a complex conjugate pair at the first (largest) caustic angle, that pair divides into two real values at the second caustic angle, the other real pair coalesces into a complex conjugate pair at the third caustic angle, and that pair finally divides into two real values at the fourth (smallest) caustic angle. In this way,

the caustics divide the ζ, ξ plane into two sectors (roughly vertically above and below the dominant region of the transient solution) in which some local wavenumber components grow exponentially, and two sectors (roughly east and west of the dominant part of the transient solution) in which all local wavenumber components propagate without growth.

At small R (Fig. 6b), the range of μ for which all stationary phase points are real is much smaller than for larger R, and correspondingly the region of the ζ, ξ plane in which all local wavenumber components propagate without growth is much smaller for radii near the central region of the solution than for radii further away.

Is the transient solution absolutely or convectively unstable? This question cannot be definitively answered using the method of Fourier summation because both limited resolution and "wrap around" make it difficult to examine the long time behavior of the solution at a fixed point. With the stationary phase solution, however, it is relatively easy to evaluate, for example, $\eta_2(x, y, t)$ either at fixed x, y as t increases or else at fixed x/t, y/t as t increases. Figure 7 shows the result at a number of different times.

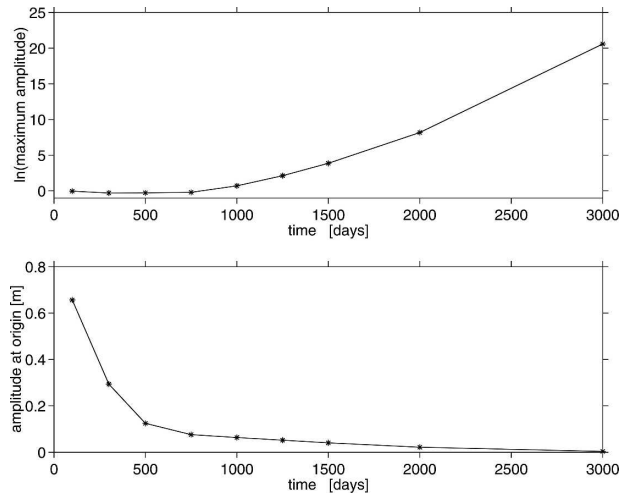The growth following the disturbance is clearly some-

FIG. 7. (top) The natural logarithm of the maximum amplitude of the middle layer interface displacement $\eta_2(x, y, t)$, moving in physical space with the disturbance, grows more rapidly than linearly as the time increases. (bottom) The amplitude of $\eta_2(x, y, t)$ at the origin decays with time. The parameters used here are the same as in Figs. 1, 2, 4, and 5.

what more rapid than a simple exponential in time. This reflects the successive instantaneous dominance of ever shorter, and hence (Fig. 1) more unstable, Fourier components. As noted above, the present solutions will no longer be valid when Fourier components whose scales approach the baroclinic Rossby radii are excited. That limit has not been reached in any of the solutions presented here but it ultimately limits their validity.

Since the solution at the origin ultimately tends toward zero even though the solution following the disturbance clearly grows without limit, we may say that the PG solution is convectively unstable, but the PG approximation precludes literally taking $t \to \infty$.

5. Discussion and conclusions

The purpose of this paper is to study the nature of long baroclinic wave propagation and instability in an idealized unventilated subtropical gyre flow with vertical shear. The new results are a detailed analytical description of the spatial structure of the long planetary wave disturbance that results when the flow in the eastern part of an unventilated steady subtropical gyre is perturbed by a region of spatially localized Ekman pumping initiated at some initial time and thereafter oscillating at a frequency on the order of a cycle per year. Both the background flow and the perturbations are solutions of a 2½-layer model, the region of forcing is located in the eastern part of the gyre where the steady flow is confined to the uppermost layer.

Besides linearization and restriction to unventilated background flows, the most important approximations are (i) neglect of relative vorticity so that only long (PG) planetary waves are considered and (ii) neglect of the spatial variation of the background flow so that phenomena such as refraction are neglected. The first approximation is not violated provided that the horizontal scale of the forcing is initially chosen to be significantly larger than the baroclinic Rossby radii characterizing the background flow. Chelton et al. (2004) showed that persistent wind features exist on scales of several hundred kilometers. The solutions studied in this paper are all for times shorter than the time at which the horizontal scales of the disturbance have decreased to the Rossby radii characterizing the background flow. The second approximation requires that the horizontal scales of the gyre be substantially greater than the horizontal scale of the forcing. Both approximations are reasonably well satisfied for long waves propagating through a background flow with weak spatial variations over sufficiently short times.

The most important results of the analysis are as follows.

1. The result of spatially localized surface Ekman pumping initiated at some initial time and thereafter oscillating at a frequency on the order of a cycle per year is (i) a directly forced part of the solution whose stable wavenumber contributions form a well-defined and narrow beam of stable waves emanating from the forcing region (Fig. 4), co-oscillating at the forcing frequency and not growing spatially plus an unstable contribution that grows spatially toward the southwest, and (ii) an accompanying transient (Figs. 2 and 5) that emanates from the forcing region with a westward component of velocity and grows in time as it propagates.

2. The direction of the co-oscillating beam is fixed (to be slightly north of westward) by the dispersion relation for free waves and by the degree of spatial localization of the forcing region (forcing regions of large lateral extent radiate much longer waves than those of small lateral extent). For the particular parameters of the examples of this paper, a Gaussian forcing region of lateral scale 500 km and frequency 1 cpy radiates strongly but, if the frequency is increased to several cycles per year the wavelengths allowed by the dispersion relation are not efficiently radiated from the forcing, and the radiation is strongly attenuated.

3. The structure of the transient is somewhat similar to that found by Kubokawa and Nagakura (2002) in the unventilated part of their subtropical gyre, but the

method of solution (stationary phase) yields a more complete description of the space and time variation of the solution—for example identifying the frontal feature (43) and the caustics that propagate westward without change of form even though the form of the full solution evolves as it propagates westward—and also (always within the limitations of the PG approximation) allows a clear determination that the transient is convectively—but not globally—unstable.

Spall (1994) made use of QG theory in a case in which the mean flow was only a function of depth to study possibly unstable perturbations of the background state representative of the real ocean, and found the fastest growing perturbations to be at scales not greatly larger than the largest baroclinic Rossby radius $O(100$ km$)$. The purpose of the present study is to provide detailed analysis of baroclinic instability acting on the larger scales at which the planetary geostrophic approximation is valid. To generalize these results to more realistic ocean subtropical gyre flows, it would be necessary to allow for the presence of smaller scales and horizontal variation of the background flow. Further analysis is thus required to determine how the horizontal spatial variation of the background flow, not allowed for in this paper, affects these solutions. It would be of particular interest to determine whether the baroclinic instability mechanism identified in this paper at larger PG scales would still play a significant role when dynamics important at scales of the order of baroclinic Rossby radius of deformation are included.

APPENDIX A

The Dynamical Model

The steady flow is modeled as in Luyten et al. (1983). Three layers (upper, middle, and lower, labeled 1, 2, and 3) are considered. The depth of the lower layer will ultimately be allowed to become very large. The momentum balance is geostrophic,

$$-fv_j = -P_{jx} \quad \text{and} \quad fu_j = -P_{jy}, \qquad \text{(A1)}$$

where subscripts x and y denote differentiation, u_j and v_j are horizontal fluid velocities in the zonal (x) and meridional (y) directions within layer j, and the Coriolis parameter f is linear in y, $f(y) = f_0 + \beta y$. $P_j = (p_j + \rho_j gz)/\rho_0$ is the Montgomery function, whose gradient depends only on the lateral position (x, y) within each layer; p_j is the true dynamic pressure in j layer, ρ_j is the (constant) density in the layer, and ρ_0 is a reference density. The vertical coordinate z is positive upward.

If H_j is the depth of the base of layer j, then the hydrostatic assumption causes the Montgomery function in layer j to be related to that in layer $j + 1$ by

$$P_j = \gamma_j H_j + P_{j+1}, \quad j = 1, 2 \qquad \text{(A2)}$$

in which $\gamma_j = g(\rho_{j+1} - \rho_j)/\rho_0$ are reduced-gravity parameters. In terms of the layer thickness $h_j = H_j - H_{j-1}$, the mass conservation equation for layer j is

$$h_{jt} + J\left(P_j, \frac{h_j}{f}\right) = \begin{cases} -w_E & \text{for} \quad j = 1 \\ 0 & \text{for} \quad j = 2, 3 \end{cases} \qquad \text{(A3)}$$

in which subscript t denotes the time derivative, J denotes the Jacobian $J(a, b) = -a_y b_x + a_x b_y$, and w_E is the wind-induced vertical velocity at the base of the surface Ekman layer. The sum of (A3) with constant H_3 gives the Sverdrup transport relation

$$-\frac{\beta}{f^2}\left(H_3 P_3 + \frac{1}{2}\gamma_1 H_1^2 + \frac{1}{2}\gamma_2 H_2^2\right)_x = -w_E. \qquad \text{(A4)}$$

The steady flow occupies a midlatitude basin of zonal extent a and meridional extent Δ with meridional boundaries at $y = b - \Delta$ and $y = b$ and is driven by Ekman pumping of amplitude W_0 having the form

$$w_E^{(0)}(y) = -W_0\left(\frac{f_0}{f}\right)^2 \sin\left[\frac{(y - b)\pi}{\Delta}\right] \qquad \text{(A5)}$$

in which f_0 is the Coriolis parameter at the central latitude of the basin. The steady Ekman pumping is chosen to be sufficiently weak that the deeper layers of the model never surface so that the steady flow occupies only the uppermost layer.

We study time-dependent perturbations linearized about the steady flow that are excited by harmonic forcing turned on at some initial time or by initial perturbations. We separate the variables into steady part and time-dependent part as

$$P_j = P_j^0 + P_j', \quad w_E = w_E^0 + w_E',$$

$$H_1 = H_1^0 - \eta_1, \quad \text{and} \quad H_2 = H_2^0 - \eta_2. \qquad \text{(A6)}$$

The perturbation mass conservation equations (A3) for the middle and the lower layer become

$$\eta_{1t} - \eta_{2t} + J\left(P_3' - \gamma_2\eta_2, \frac{H_{20} - H_1^0}{f} \right) = 0 \quad (A7)$$

and

$$\eta_{2t} + J\left(P_3', \frac{H_{30} - H_{20}}{f} \right) = 0 \quad (A8)$$

where $H_2^0 = H_{20}$ is the depth of the bottom of the middle layer (constant when the steady flow in that layer is stagnant) and H_{30} is the flat bottom of the basin. The perturbation Sverdrup equation, obtained by integrating (A4) westward from the eastern boundary $x = a$ and linearizing about the steady flow, is

$$-\frac{\beta}{f^2}[H_{30}P_{3x}' - \gamma_1(H_1^0\eta_1)_x - \gamma_2 H_{20}\eta_{2x}] = -w_E.$$

$$(A9)$$

We next reduce the model to two layers lying over an infinitely deep quiescent third layer to obtain the so-called 2½-layer model. From (A8), $P_3' \to 0$ as $H_{30} \to \infty$ so that (A7) becomes

$$\eta_{1t} - \eta_{2t} - J\left(\eta_2, \gamma_2 \frac{H_{20} - H_1^0}{f} \right) = 0. \quad (A10)$$

We then solve the Sverdrup equation (A9) for P_3', insert into (A8), and let the total depth $H_{30} \to \infty$, to obtain

$$\eta_{2t} - \frac{\beta\gamma_2 H_{20}}{f^2}\eta_{2x} - \frac{\beta\gamma_1}{f^2}(H_1^0\eta_1)_x = w_E. \quad (A11)$$

Equations (A10) and (A11) constitute the 2½-layer model of the time-dependent perturbations η_1 and η_2 of nonzonal background flow.

APPENDIX B

Stationary Phase Points

To determine the set of stationary phase points in the complex α plane over which to sum in (40), it is necessary to deform the integration path $\alpha = (0, 2\pi)$ of (32) into an integration path that passes through each stationary phase point along that point's path of steepest descent. This must be done separately for the integrands involving σ^+ and those involving σ^-.

The only singularities of the integrands of the terms (32) in the complex α plane are at the simple zeros of H^2 (36), whose square root appears in the phase (33). These zeros occur in real pairs, the location of the second pair being the location of the first pair plus π radians. In deforming the α contour of integration the two branch cuts joining the members of each pair separately

must not be crossed. It is convenient to choose the branch cuts to lie along the real axis between the two members of each individual pair.

Once H has been determined, solution of (36) for the stationary phase values of α yields four stationary phase points whose real parts lie between $(0, \pi)$ and another four stationary phase points whose real parts in $(\pi, 2\pi)$ are the real parts of the first four plus π. For every physical space point (ξ, ζ), the interval $(0, \pi)$ contains three stationary phase points (either three real points or one real point and one complex conjugate pair) associated with σ^+ and one (real) point associated with σ^- or vice versa. A similar statement holds for the stationary phase points whose real parts lie in $(\pi, 2\pi)$.

When all stationary phase points are real, no deformation of the integration path $\alpha = (0, 2\pi)$ is needed, and the sum in (40) must include contributions from all stationary phase points. When the integrand of (32) under consideration has one real stationary phase point and a complex conjugate pair whose real parts is in $(0, \pi)$ with corresponding points in $(\pi, 2\pi)$, it is not possible to deform the contour $\alpha = (0, 2\pi)$ to lie entirely along the steepest descent paths threading one or more of the stationary phase points without crossing the branch cuts joining the simple zeros of H^2 (36). If however the deformed contour is made to parallel the branch cuts where it meets them, then along those cuts the imaginary part of the phase in the integrand is positive and large so that as $t \to \infty$ the contributions to the integral from these portions of the deformed path will be exponentially smaller than those from the stationary phase points. Because the complex stationary phase points come in complex conjugate pairs, the sum over all stationary phase points threaded by the deformed contour is the same as the sum over all the real stationary phase points plus one-half times the sum over all the complex stationary phase points. This then is the general rule for each integrand of (32).

APPENDIX C

Approximate Stationary Phase Solutions Valid at Large Distances away from the Center of the Disturbance

We seek approximate analytic solutions, valid at large distances away from the center of the disturbance, of the quartic equation (38) for H as defined in (34), which yield the four values of the stationary phase angle α_0.

Introduce the nondimensional variables as

$$H = |v|^{1/2}H', \quad b = |v|b', \quad \eta = |v|^{1/2}\eta', \quad \text{and} \quad \xi = |v|^{1/2}\xi'$$

$$(C1)$$

and transform to cylindrical polar coordinates to write

$$\xi = R\cos\mu \quad \text{and} \quad \eta = R\sin\mu. \tag{C2}$$

In terms of (C1) and (C2) (where we have dropped the primes for the convenience) (38) becomes

$$H[H^2 - (b + v\tan\mu)] + \frac{H^4 - (b^2 + 1)}{-2R\cos\mu} = 0. \tag{C3}$$

We seek approximate solutions of this equation for large R.

First seek $O(1)$ solution by neglecting much smaller terms of the order R^{-1}; one iteration on the two roots thus obtained yields

$$H_{1,2} = \pm\left[b + v\tan\mu + \frac{(b + v\tan\mu)^2 - (b^2 + 1)}{\pm 2R\cos\mu(b + v\tan\mu)^{1/2}} \right]^{1/2}. \tag{C4}$$

From the dispersion relation (36)

$$\tan\alpha_{01,2} = \tan\mu + \frac{1}{v}\left[\frac{(b + v\tan\mu)^2 - (b^2 + 1)}{\pm 2R\cos\mu(b + v\tan\mu)^{1/2}} \right], \tag{C5}$$

and the two values of α_0 are given by

$$\alpha_{01,2} = \mu \pm O(R^{-1}). \tag{C6}$$

They are two $O(1)$ solutions for which α_0 is close to μ, but the correction term in (C5) diverges at $\mu = \pi/2$ and $\mu = \arctan(-b/v)$.

An $O(R)$ solution is obtained from the two highest-order terms in (C3), which are $O(R^3)$, and give $H^4 - 2H^3R\cos\mu = 0$, whose solution is $H = 2R\cos\mu$. In terms of the angle μ_0 defined by

$$\sin\mu_0 = \frac{b}{\sqrt{b^2 + 1}} \quad \text{and} \quad \cos\mu_0 = -\frac{v}{\sqrt{b^2 + 1}}, \tag{C7}$$

the once iterated $O(R)$ solution becomes

$$H_3 = 2R\cos\mu + \frac{\sqrt{b^2 + 1}}{2R\cos^2\mu}\sin(\mu - \mu_0) \tag{C8}$$

and yields α_0:

$$\alpha_{03} = \frac{\pi}{2}. \tag{C9}$$

The correction of this root blows up at $\mu = \pi/2$.

An $O(R^{-1})$ solution is approximated by

$$H_4 = \frac{\sqrt{b^2 + 1}}{2R\sin(\mu_0 - \mu)}. \tag{C10}$$

It becomes, after one iterative correction,

$$H_4 = -\frac{\sqrt{b^2 + 1}}{2R\sin(\mu - \mu_0)} + \frac{(b^2 + 1)^{3/2}\cos\mu}{(2R)^3\sin^4(\mu_0 - \mu)} \tag{C11}$$

and yields α_0

$$\alpha_{04} = \mu_0. \tag{C12}$$

This correction blows up for $\mu = \mu_0$. The four roots of α_0 given by (C6), (C9), and (C12) are the approximations to the angle which makes the phase stationary and over which (40) must be summed. When they are inserted in (40), the result shows that the asymptotic solutions for large R decay transcendentally away from the center of the front. Solutions (C6), (C9), and (C12) are crude approximations, and the correction terms diverge at $\mu = \pi/2$ and $\mu = \mu_0 = \arctan(-b/v)$. The more accurate expressions for the position of the caustics at large distances from the center of the disturbance, as well as approximate solutions for large distances that are uniformly valid across the caustics, can be obtained by keeping all the terms of the quartic equation in (38) that interact at the caustic considered.

REFERENCES

Barnier, B., 1988: A numerical study on the influence of the Mid-Atlantic ridge on non linear first-mode baroclinic Rossby waves generated by seasonal winds. *J. Phys. Oceanogr.,* **18,** 417–433.

Briggs, R. J., 1964: *Electron–Stream Interaction with Plasmas.* Special Tech. Rep. 10, *MIT Press Research Monogr.,* MIT Press, 183 pp.

Carrier, G. F., M. Krook, and C. E. Pearson, 1966: *Functions of a Complex Variable: Theory and Technique.* McGraw-Hill, 438 pp.

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

——, M. Schlax, M. Freilich, and R. Milliff, 2004: Satellite measurements reveal persistent small-scale features in ocean winds. *Science,* **303,** 978–983.

de Szoeke, R. A., and D. Chelton, 1999: The modification of long planetary waves by homogeneous potential vorticity layers. *J. Phys. Oceanogr.,* **29,** 500–511.

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

——, 1998: On "too-fast" baroclinic planetary waves in general circulation. *J. Phys. Oceanogr.,* **28,** 1739–1758.

——, and R. X. Huang, 2001: Adjustment of the ventilated thermocline. *J. Phys. Oceanogr.,* **31,** 1676–1697.

Huerre, P., and P. A. Monkewitz, 1990: Local and global instabilities in spatially developing flows. *Annu. Rev. Fluid Mech.,* **22,** 473–537.

Kubokawa, A., and M. Nagakura, 2002: Linear planetary wave dynamics in a 2.5-layer ventilated thermocline model. *J. Mar. Res.,* **60,** 367–404.

Lighthill, J. M., 1978: *Waves in Fluids.* Cambridge University Press, 504 pp.

Liu, Z., 1999a: Planetary wave modes in the thermocline: Non-Doppler-shift mode, advective mode and Green mode. *Quart. J. Roy. Meteor. Soc.,* **125,** 1315–1339.

——, 1999b: Forced planetary wave response in a thermocline gyre. *J. Phys. Oceanogr.,* **29,** 1036–1055.

Luyten, J. R., J. Pedlosky, and H. Stommel, 1983: The ventilated thermocline. *J. Phys. Oceanogr.,* **13,** 828–857.

Pedlosky, J., 1987: *Geophysical Fluid Dynamics.* 2d ed. Springer-Verlag, 710 pp.

Spall, M. A., 1994: Mechanisms for low-frequency variability and salt flux in the Mediterranean salt tongue. *J. Geophys. Res.,* **99,** 10 121–10 129.

Tailleux, R., and J. McWilliams, 2002: Energy propagation of long extratropical Rossby waves over slowly varying zonal topography. *J. Fluid Mech.,* **473,** 295–319.

Walker, A., and J. Pedlosky, 2002: Instability of meridional baroclinic currents. *J. Phys. Oceanogr.,* **32,** 1075–1093.

Wang, L., C. J. Koblinsky, and S. Howden, 2001: Annual Rossby wave in the southern Indian Ocean: Why does it "appear" to break down in the middle ocean? *J. Phys. Oceanogr.,* **31,** 54–75.