Finite-Amplitude Baroclinic Instability of Time-Varying Abyssal Currents

SEUNG-JI HA AND GORDON E. SWATERS

Applied Mathematics Institute, Department of Mathematical and Statistical Sciences, and Institute for Geophysical Research, University of Alberta, Edmonton, Alberta, Canada

ABSTRACT

The weakly nonlinear baroclinic instability characteristics of time-varying grounded abyssal flow on sloping topography with dissipation are described. Specifically, the finite-amplitude evolution of marginally unstable or stable abyssal flow both at and removed from the point of marginal stability (i.e., the minimum shear required for instability) is determined. The equations governing the evolution of time-varying dissipative abyssal flow not at the point of marginal stability are identical to those previously obtained for the Phillips model for zonal flow on a β plane. The stability problem at the point of marginally stability is fully nonlinear at leading order. A wave packet model is introduced to examine the role of dissipation and time variability in the background abyssal current. This model is a generalization of one introduced for the baroclinic instability of zonal flow on a β plane. A spectral decomposition and truncation leads, in the absence of time variability in the background flow and dissipation, to the *sine–Gordon* solitary wave equation that has grounded abyssal soliton solutions. The modulation characteristics of the soliton are determined when the underlying abyssal current is marginally stable or unstable and possesses time variability and/or dissipation. The theory is illustrated with examples.

---

1. Introduction

Recent work on the baroclinic instability of time-varying flow has suggested that time dependence can have a profound effect on the stability properties of ocean currents. For example, Pedlosky and Thomson (2003, hereinafter referred to as PT), in a study of the two-layer Phillips model of baroclinic instability of a zonal flow on a β plane, have shown that simple time variations in the zonal current can destabilize the flow even if the time average of the current is itself stable (and vice versa). Such time dependence occurs on many different time scales and for many different reasons. Some of these reasons include tidally forced flow variations, weather-system-induced variability, seasonal variations, or even longer time scale interannual variability. In the context of source-driven abyssal flow it is easy to imagine that there are seasonal variations in the

intensity of the atmospheric cooling that produce the deep convection and this, in turn, will result in a time-varying abyssal current.

The principal purpose of this paper is to develop a nonlinear theory for marginally stable or unstable, time-varying grounded abyssal currents. Swaters (1991) described the linear baroclinic instability of grounded abyssal currents on a sloping bottom. The instability mechanism modeled by Swaters (1991) is the release of the available gravitational potential energy (AGPE) associated with a dense water mass sitting directly, that is, *grounded*, on a sloping bottom surrounded by relatively lighter water. As the abyssal current becomes unstable, downslope-propagating plumes develop on the offshore isopycnal incropping or grounding (see Fig. 1). The AGPE is transferred to perturbation kinetic energy in the overlying water column that is organized into topographic Rossby waves. Jiang and Garwood (1996), Jungclaus et al. (2001), Etling et al. (2000), and others have concluded that the instabilities observed in three-dimensional numerical simulations of overflows on a continental slope arise from Swaters's (1991) instability mechanism.

Mooney and Swaters (1996) developed a finite-

Corresponding author address: Gordon E. Swaters, Applied Mathematics Institute, Dept. of Mathematical and Statistical Sciences, and Institute for Geophysical Research, University of Alberta, Edmonton, AB T6G 2G1, Canada.
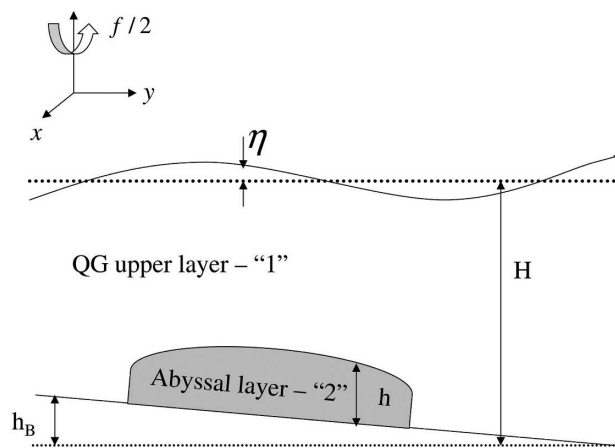E-mail: gordon.swaters@ualberta.ca

FIG. 1. Geometry of the model used in this paper. The upper and abyssal layers are governed by quasigeostrophic and planetary geostrophic dynamics, respectively. The abyssal-layer height, h, can have groundings or incroppings, i.e., distinct locations where $h = 0$.

amplitude instability theory (see, also, Swaters 1993) for abyssal currents based on the Swaters's (1991) model. They showed that it was possible for the instabilities to saturate and for the unstable abyssal current to evolve toward a new quasi-equilibrium state or to form solitary wave packets. Numerical simulations by Swaters (1998a) showed that predictions of the weakly nonlinear Mooney and Swaters (1996) theory remained true even in the fully nonlinear regime. Swaters's (1998a) simulations showed that it was possible for along-slope-propagating grounded abyssal domes to emerge from the instability process (see, also, Swaters 1998b).

Poulin and Swaters (1999a,b,c) extended the Swaters (1991) model to the situation where the overlying water column is continuously stratified. Subsequently, Reszka et al. (2002) developed the linear instability theory for this model and presented numerical simulations for the nonlinear evolution of source-driven abyssal flows for parameter values characteristic of the Denmark Strait overflow (DSO). In addition to showing how this model could reproduce the spatial and temporal characteristics of the mesoscale variability observed in the DSO, these simulations were also able to reproduce the formation of surface-intensified eddies that have been observed in satellite imagery (Bruce 1995).

Pavec et al. (2005) have recently described, following the ideas in PT, the *linear* baroclinic instability of marginally unstable inviscid oscillatory abyssal currents *not* at the "point of marginal stability" [i.e., in the present context, the minimum baroclinic shear required for instability; see, e.g., Drazin and Reid (1981)]. Here, the work of Pavec et al. (2005) is extended into the nonlin-

ear regime for *both* marginally stable and unstable abyssal flows with general long time scale variability with dissipation. In addition, the nonlinear theory is developed for marginally stable or unstable time-varying abyssal flows that *are* at the point of marginal stability. This latter situation is the most physically relevant "transition to instability" problem.

As is shown by Mooney and Swaters (1996) for abyssal currents, but is well known in the nonlinear theory of the baroclinic instability of a zonal flow on a β plane in the context of the Phillips model (Pedlosky 1970, 1972, 1982a,b), the nonlinear development of marginally unstable flows depends crucially on the underlying perturbation mode and flow configuration being examined. When the perturbation mode *does not* correspond to the point of marginal stability, there is an amplitude separation between the "fundamental mode" and the "harmonics" that nonlinearity generates. That is, the transition to instability is described by a sequence of linear equations. However, when the perturbation mode *does* correspond to the point of marginal stability, then even at lowest order, the disturbance field is fully nonlinear [this was first shown for the Phillips model for the baroclinic instability of a steady zonal flow by Pedlosky (1982a,b), and later was more completely discussed by Warn and Gauthier (1989)] and it is impossible to introduce an amplitude separation between the fundamental and its harmonics.

The plan of this paper is as follows. In section 2, the principal stability properties of the Swaters (1991) baroclinic model for grounded abyssal flow are reviewed.

In section 3, the finite-amplitude theory for the baroclinic instability of time-varying abyssal flow with dissipation present that *does not* correspond to the point of marginal stability is developed. It is shown that the coupled pair of amplitude equations that is derived is identical (modulo a trivial rescaling) to the model obtained and extensively analyzed by PT. A brief review of the relevant results is given.

In section 4, the finite-amplitude baroclinic wave packet instability theory for time-varying abyssal flow, with dissipation present, that *does* correspond to the point of marginal stability is developed. Wave packet equations are derived that are the analog of the Warn and Gauthier (1989) model for marginally unstable baroclinic flow. A spectral decomposition is introduced (Mooney and Swaters 1996) that leads to an infinite set of coupled spatial–temporal amplitude equations. These equations are intractable. However, if an ad hoc truncation is introduced that retains only the fundamental mode and the mean flow it generates, then the resulting model equations are equivalent to the sine–

Gordon equation (Gibbon et al. 1979) that has a soliton solution.

The truncated model has been referred to as the *AB equations* (see, e.g., Tan and Boyd 2002). The history of the AB equations is a nice illustration of the interplay between numerical simulations and theory. Pedlosky (1972) first derived the AB model for a constant potential vorticity flow at minimum baroclinic shear. As described above, a resonance develops between the fundamental and all other harmonics leading to the nonseparability of the modes (i.e., a nonlinear critical layer develops everywhere in the flow). This property was first clearly seen in the numerical simulations of Boville (1981), which then lead to the development of the infinite-dimensional modal theory by Pedlosky (1982a). However, as is also shown by Pedlosky (1982a), the development of the nonlinear critical layer everywhere in the flow *does not occur* in the generic, and more physically realistic, situation where the background flow does not have *constant potential vorticity*. In this case, an amplitude separation between the fundamental and the higher harmonics occurs and the resulting finite-amplitude theory results in the *modified* AB equations (possessing a single additional cubic term). As commented on by Tan and Boyd (2002), the striking difference between the weakly nonlinear models for mean flows that possess no spatial shear and those that do is common in singular perturbation theory.

Irrespective of these comments, it is appropriate to acknowledge the simplification made here. Explicitly, and to quote Tan and Boyd (2002): "the classical *AB* system is *not* a consistent model of baroclinic waves at marginal shear. It can be justified as two-latitudinal mode truncation of the correct system for marginally unstable waves on uniform potential vorticity. It can be also justified as the proper system for more general mean flows when the cubically nonlinear term is negligible." It is our view that, notwithstanding the simplification, the results obtained from the truncated model will provide important "guide posts" for the analysis of solutions to the extended Warn and Gauthier (1989) equations.

A nonlinear Wentzel–Kramers–Brillouin (WKB) technique is introduced to determine how the soliton solution to the truncated equations is modulated if the background marginal abyssal flow is time varying with dissipation present. As is known based on inverse scattering theory for perturbed solitons (Kaup and Newell 1978), the evolution of the soliton parameters will satisfy an averaged energy balance equation. The averaged energy balance results in a time-dependent differential "transport" equation that determines the evolution of the soliton translation velocity, amplitude, and "wavenumber." Detailed descriptions of the evolution of the soliton parameters are given when dissipation is present (but without time variability in the background marginal flow) and when time variability is present (but without dissipation). Because the underlying mathematical theory associated with the inverse scattering transform (Ablowitz and Segur 1981) implies smooth dependence on the initial data, the evolution of the perturbed soliton with *both* dissipation and time variability present will vary continuously between these limiting solutions.

The finite interval of allowed soliton translation velocities associated with a marginally *stable* background flow separates the allowed set of soliton translation velocities associated with a marginally *unstable* background flow into two disconnected semi-infinite intervals (Pedlosky 1972; Gibbon et al. 1979). The two values for the soliton translation velocity that separate the unstable and stable regions correspond to, respectively, the zero solution and a singular limit that can only exist if the background flow has no baroclinic shear.

It is shown that the dissipating soliton always evolves toward the zero solution. However, because of the structure of the solutions associated with the boundaries, with respect to the soliton translation velocity, between the unstable and stable regions, complex behavior can develop as the soliton is modulated as a result of dissipation or time variability. The structure and behavior of the solutions is fully described. In particular, a solution to the transport equation is identified that can connect the marginally stable and unstable solitons if time variability in the background shear goes from being sub- to supercritical (and vice versa), that is, if the background shear passes through zero. These properties are illustrated with examples.

The first-order perturbation equations are examined when time variability and dissipation are present. The averaged energy balance is equivalent to a solvability condition on these equations. It is shown that in satisfying the averaged energy balance, the slowly deforming soliton is unable to simultaneously satisfy the averaged mass balance. This results in the emergence of a small-amplitude "shelf region" (Knickerbocker and Newell 1980; Kodama and Ablowitz 1980) in the nonlinearly induced mean flow that arises behind the disturbed propagating soliton. Last, the leading-order structure of the perturbation field associated with the modulated soliton, when dissipation and time variability are present, is determined.

In section 5, conclusions and suggestions for future work are given.

2. The governing equations

The *nondimensional* Swaters (1991) model equations, with dissipation proportional to the potential vorticity (PV), can be written in the form

$$\Delta\eta_t - J(h_B, h) + J(\eta, \Delta\eta + h_B) = -r\Delta\eta \quad \text{and} \qquad (2.1\text{a})$$

$$h_t + J(\eta + h_B, h) = -rh, \qquad (2.1\text{b})$$

where η, h, and h_B are the upper-layer geostrophic pressure, abyssal-layer height, and height of the bottom topography above the mean depth (see Fig. 1), respectively, with the auxiliary relations

$$\mathbf{u}_1 = \hat{\mathbf{e}}_3 \times \nabla\eta, \mathbf{u}_2 = \hat{\mathbf{e}}_3 \times \nabla(h_B + \eta + h), \quad \text{and}$$

$$p = \eta + h - y, \qquad (2.2)$$

where $\mathbf{u}_1$, $\mathbf{u}_2$, and p are the upper- and abyssal-layer velocities, and the abyssal-layer geostrophic pressure, respectively. The notation is standard and x and y are the along- and cross-slope coordinates, respectively. Henceforth, it will be assumed that $h_B = -y$ [the bottom topography is scaled using its slope; see Swaters (1991)]. The spatial domain is the channel $|x| < \infty$ and $0 < y < L$. The boundary conditions are $v_1 = v_2 = 0$ on $y = 0, L$. The sum (2.1a) + (2.1b) is the quasigeostrophic (QG) PV equation and $-(2.1\text{b})/h^2$ is the planetary geostrophic (PG) PV equation for the upper and abyssal layers, respectively.

The model (2.1) can be formally derived in a small Rossby number limit of the two-layer shallow water equations (Swaters 1991). Briefly, the abyssal-layer equations are scaled assuming that the dynamics is principally governed by a geostrophic balance between the downslope gravitational acceleration of the abyssal water mass and the Coriolis term. The upper-layer dynamics is scaled assuming that the baroclinic stretching associated with deformations of the interface between the abyssal and upper layer is the same order of magnitude as the relative vorticity field. This will imply that the appropriate length scale is the internal deformation radius associated with the upper layer and that time will be scaled advectively. All other variables are scaled assuming an underlying geostrophic balance to leading order.

From the point of view of the abyssal layer, this is an "intermediate," or PG, dynamical limit [i.e., a subinertial regime in which the length scale is longer than the local internal deformation radius *associated with the abyssal layer* but shorter than the basin width; see Charney and Flierl (1981), Flierl (1984), and Pedlosky (1984)]. The dominant nonlinearity is associated *not* with the flow acceleration but, rather, with isopycnal steepening. This attribute allows the model to describe fully grounded abyssal flow, in which the isopycnal field intersects the bottom (see Fig. 1), which is something QG theory cannot do. This scaling allows for strong baroclinic interaction between the abyssal current and the overlying water column.

Implicit in the derivation of (2.1) is that the ratio of the scale thickness of the abyssal layer to the scale thickness of the overlying ocean is the same order of magnitude as variations in the topographic height over the internal deformation radius associated with the overlying ocean. This implies that the overlying ocean will have a finite deformation radius and, thus, baroclinic stretching associated with the deforming interface between the upper and abyssal layer and a background topographic vorticity gradient is retained in the upper-layer PV balance. This fact has a very important implication in relation to the stability properties of the model. It is known that, in the purely inertial limit, the PG approximation exhibits an ultraviolet catastrophe in the linear instability problem [i.e., the most unstable occurs for an infinite wavenumber; see de Verdiere (1986)]. While the inclusion of Rayleigh damping can remove the ultraviolet catastrophe (Samelson and Vallis 1997), Swaters (1991) has shown that the inviscid $r = 0$ limit in (2.1), which couples an abyssal PG layer to an overlying QG layer (with its implicit finite deformation radius), also ensures that the most unstable mode occurs at a finite wavenumber.

Linear stability equations and properties

To derive the stability equations, the decomposition

$$h = h_0(y) + h'(x, y, t), \eta = \eta'(x, y, t), \qquad (2.3)$$

is introduced, where $h_0(y)$ is the steady abyssal current height that is an exact solution to the *inviscid* model (in order to focus on the baroclinic instability of abyssal currents, it is assumed that there is no mean flow in the upper layer), and η' and h' are the perturbations. Substitution into (2.1a,b) yields (after dropping the primes) the *nonlinear perturbation equations:*

$$\Delta\eta_t - \eta_x - h_x + J(\eta, \Delta\eta) = -r\Delta\eta \quad \text{and}$$

$$(\partial_t + \partial_x)h + h_{0y}\eta_x + J(\eta, h) = -rh. \qquad (2.4)$$

These will be used in the weakly nonlinear analysis presented in the next two sections.

It is useful to briefly review the inviscid linear instability theory. The inviscid ($r = 0$) *linear stability equations* are given by

$$\Delta\eta_t - \eta_x - h_x = 0 \quad \text{and}$$

$$(\partial_t + \partial_x)h + h_{0y}\eta_x = 0. \qquad (2.5)$$

The *normal mode linear instability equations*, obtained by substituting

$$(\eta, h) = [\hat{\eta}(y), \hat{h}(y)] \exp[ik(x - ct)] + \text{c.c.}, \quad (2.6)$$

into (2.5), where c.c. means the complex conjugate of the preceding term, k is the real-valued along-slope wavenumber, and $c = c_R + ic_I$ is the along-slope complex-valued phase velocity, are

$$\hat{\eta}_{yy} - \left[k^2 - \frac{1}{c} - \frac{h_{0y}}{c(c-1)}\right]\hat{\eta} = 0 \quad \text{and} \quad \hat{h} = \frac{h_{0y}}{c-1}\hat{\eta}, \quad (2.7)$$

with the boundary conditions $\hat{\eta} = \hat{h} = 0$ on $y = 0, L$.

In the remainder of this paper attention will be restricted to the constant abyssal flow given by

$$h_0(y) = h_{\max} - \gamma y, \gamma > 0, \quad (2.8)$$

where it is assumed that $h_{\max} - \gamma y > 0$ for $y \in (0, L)$. This profile is the same one used by Mooney and Swaters (1996) and since it implies a constant, that is, horizontally unsheared, abyssal current velocity, it is similar to that assumed in PT.

Substituting (2.8) into the normal-mode equations in (2.7) yields

$$\hat{\eta}_{yy} - \left[k^2 - \frac{1}{c} + \frac{\gamma}{c(c-1)}\right]\hat{\eta} = 0 \quad \text{and} \quad \hat{h} = -\frac{\gamma}{c-1}\hat{\eta}. \quad (2.9)$$

The solution to (2.9) is given by

$$(\eta, h) = A\left(1, \frac{\gamma}{1-c}\right) \sin(ly) \exp[ik(x - ct)] + \text{c.c.}, \quad (2.10)$$

where A is a free constant and $l = n\pi/L$ with $n \in \mathbb{Z}^+$, with the dispersion relation

$$c = \frac{K^2 + 1 \pm [(K^2 - 1)^2 - 4K^2\gamma]^{1/2}}{2K^2}, \quad (2.11)$$

where $K = \sqrt{k^2 + l^2}$ is the total wavenumber.

The boundary between stability and instability is determined by the quantity within the square root in (2.11) being zero. This yields the marginal stability curve

$$\gamma = \gamma_c(K) \equiv (K^2 - 1)^2/(4K^2) \geq 0.$$

A mode with total wavenumber K is unstable *if and only if* $\gamma > \gamma_c$.

The fact that $\gamma_c > 0$ implies that only abyssal flows with $u_2 > 0$ can be baroclinically unstable [see (2.2), (2.3), and (2.8) and Fig. 1]. Since the sloping bottom
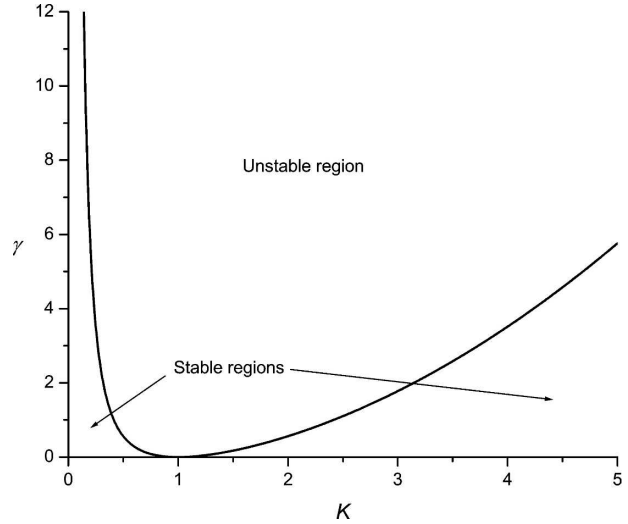

FIG. 2. The marginal stability curve $\gamma = \gamma_c(K) = (K^2 - 1)^2/4K^2$. The point of marginal stability is located at $K = 1$ and $\gamma = 0$.

acts like a topographic β plane in the *upper-layer* PV equation in (2.1a), it follows that only "westward" abyssal currents can be baroclinically unstable according to this QG–PG model. The addition of a constant flow (of either sign) in the upper layer does not change this observation since, mathematically, it can be always set equal to zero by the introduction of an appropriate Galilean transformation in (2.1a,b). That only westward abyssal flow is unstable is another reason why this model is qualitatively different than the Phillips model, in which baroclinic instability is possible for both eastward or westward flow provided the absolute vertical shear is large enough (Pedlosky 1987). The underlying reason for the difference is that in (2.1b) there is an actual body force present (i.e., the geostrophically balanced downslope acceleration of a dense water mass sitting directly, or grounded, on a sloping bottom) that gives rise to the steady along-slope westward abyssal current.

In Fig. 2 the marginal stability curve is presented. The minimum of the marginal stability curve is located at $K = 1$ and corresponds to $\gamma_c = 0$. The *point of marginal stability* corresponds to the first value of γ for which any larger value of γ leads to instability for some K. The point of marginal instability therefore corresponds to $\gamma_c = 0$ and $K = 1$. Clearly, the $K = 1$ mode can only exist if $l \leq 1$ for some value of n. Consequently, henceforth, it will be assumed that $n = 1$ so that $l = \pi/L$ (the gravest, or first, cross-slope mode).

3. Weakly nonlinear evolution of $K \neq 1$ modes

Here, the finite-amplitude theory is developed for modes that do not correspond to the point of marginal

stability, that is, a mode not located at the minimum of the marginal stability curve. To determine the proper scalings the dispersion relation (2.11) is examined in the situation when γ is slightly supercritical. Assuming that $\gamma = \gamma_c + \delta$, where δ is a small positive number, allows the dispersion relation to be written in the form

$$c = \frac{K^2 + 1}{2K^2} \pm i \frac{\sqrt{\delta}}{K}. \tag{3.1}$$

Thus, the linear growth rate will be proportional to $\sqrt{\delta}$ so that the e-folding time scale associated with the slightly supercritical mode will be $O(1/\sqrt{\delta})$. Following and extending the ideas of Mooney and Swaters (1996) and PT to allow for time variations in the slightly sub- or supercritical abyssal current, γ is chosen to be of the form

$$\gamma = \gamma_c + \varepsilon^2[\Upsilon_0 + \Upsilon(\varepsilon t)], \tag{3.2}$$

where $\Upsilon_0 = \pm 1$ [$\Upsilon_0 = +1$ and $\Upsilon_0 = -1$ correspond to a slightly supercritical (unstable) or subcritical (stable) abyssal current, respectively] and $\Upsilon(\varepsilon t)$ is an $O(1)$ real-valued function of time that models the temporal variations of the marginally stable or unstable abyssal flow. A specific form for $\Upsilon(\varepsilon t)$ will be chosen later. The parameter ε is assumed to satisfy $0 < \varepsilon \ll 1$ and corresponds to the nondimensional order of magnitude of the amplitude of the perturbation. In addition, the weakly dissipative limit $r = \varepsilon\mu$ with $\mu \simeq O(1)$ is introduced.

It is important to point out that if (3.2) is substituted into (2.4), the resulting h_0 is no longer an exact solution of the inviscid limit of (2.1b). In order that the resulting $h_0(y, \varepsilon t)$ be a solution of the governing equations, it is necessary to introduce a "forcing" term on the right-hand side of (2.1b) given by $\mathcal{F} = h_{0t} + rh_0$. The forcing term does not appear in the perturbation equations in (2.4), however, and it is not necessary to fully describe the forces that give rise to such an h_0 in order to give a complete description of its stability characteristics (see, e.g., Pedlosky's (1987, section 7.13) discussion of the baroclinic instability of nonzonal flow).

Under these assumptions, the weakly nonlinear and slow time scalings,

$$(\eta, h)(x, y, t) = \varepsilon(\tilde{\eta}, \tilde{h})(x, y, t, T; \varepsilon), \quad T = \varepsilon t, \tag{3.3}$$

are introduced, where $\tilde{\eta}$ and $\tilde{h}$ are both assumed to be $O(1)$. Substitution of (3.2) and (3.3) into (2.4), dropping the tildes, yields

$$\Delta\eta_t - \eta_x - h_x = -\varepsilon[(\partial_T + \mu)\Delta\eta + J(\eta, \Delta\eta)] \quad \text{and} \tag{3.4}$$

$$h_t + h_x - \gamma_c\eta_x = -\varepsilon[(\partial_T + \mu)h + J(\eta, h)]$$
$$+ \varepsilon^2[\Upsilon_0 + \Upsilon(T)]\eta_x. \tag{3.5}$$

These equations can be solved with an expansion of the form

$$(\eta, h)(x, y, t, T; \varepsilon) = (\eta_0, h_0)(x, y, t, T)$$
$$+ \varepsilon(\eta_1, h_1)(x, y, t, T) + \cdots \cdots \tag{3.6}$$

Substitution of (3.6) into (3.4) and (3.5) leads to the $O(1)$ solution

$$(\eta_0, h_0) = A(T)[1, (K^2 - 1)/2] \sin(ly) \exp[ik(x - ct)] + \text{c.c.}$$

and

$$c = (K^2 + 1)/(2K^2). \tag{3.7}$$

The $O(\varepsilon)$ solution is given by $\eta_1 = 0$ with h_1 given by [see Mooney and Swaters (1996) or Ha (2004) for complete details]

$$h_1(x, y, t, T) = -K^2l \sin(2ly)\phi(T) + \{iK^2k^{-1}(\partial_T + \mu)$$
$$\times A \sin(ly) \exp[ik(x - ct)] + \text{c.c.}\}, \tag{3.8}$$

where $\phi(T)$ is the time-varying part of the mean flow that is generated by the perturbation field interacting with itself [and is determined from the $O(\varepsilon^2)$ problem].

Last, from the $O(\varepsilon^2)$ problem, the removal of terms that lead to secular growth yields the pair of equations

$$(\partial_T + \mu)^2 A - \sigma^2[\Upsilon_0 + \Upsilon(T)]A + NA\phi = 0 \quad \text{and} \tag{3.9}$$

$$(\partial_T + \mu)\phi = (\partial_T + 2\mu)|A|^2, \tag{3.10}$$

where $\sigma \equiv k/K$ and $N \equiv k^2/l^2$. Note that σ is the linear growth rate for the slightly unstable mode when $\Upsilon_0 = +1$ and will be the frequency for the marginally stable mode when $\Upsilon_0 = -1$. This is in precise agreement with the linear theory as seen in (3.1).

The solution for $\phi(T)$ can be written in the form

$$\phi(T) = |A|^2(T) + e^{-\mu T}$$
$$\times \left[\mu \int_0^T e^{\mu\xi}|A|^2(\xi) \, d\xi - |A_0|^2\right], \tag{3.11}$$

where $A_0 = A(0)$ and it is assumed that $\phi(0) = 0$. Thus, (3.9) can be written in the form

$$\{(\partial_T + \mu)^2 - \sigma^2[\Upsilon_0 + \Upsilon(T)]\}A + NA$$

$$\times \left\{|A|^2(T) + e^{-\mu T}\left[\mu \int_0^T e^{\mu\xi}|A|^2(\xi)\,d\xi - |A_0|^2\right]\right\} = 0,$$

$$(3.12)$$

which will determine $A(T)$, subject to $A(0) = A_0$ and $A_T(0) = \sigma\sqrt{|\Upsilon_0|}A_0$.

The coupled pair of amplitude equations in (3.9) and (3.10) is identical (modulo a trivial rescaling) to the model obtained and extensively analyzed by PT. [See PT's (3.13a,b). However, there is a typographical error in PT's (3.13a). The second term in PT's (3.13a) is linear in B, not quadratic as written.] It is useful to briefly summarize the main results. For a comprehensive investigation of the solutions to (3.9) and (3.10) for various parameter regimes, see PT or Ha (2004).

If the underlying flow does not possess time variability [i.e., $\Upsilon(T) = 0$], (3.9) and (3.10), or equivalently, (3.12), are a Lorenz dynamical system in which the Prandtl number is one (Klein and Pedlosky 1992). With dissipation present (i.e., $\mu > 0$), the solutions always approach a steady state [i.e., $A(T) \to A_\infty$ as $T \to \infty$] and there are no chaotic or periodic solutions. Indeed, this property is the reason for choosing the dissipation in (2.1) as proportional to the dynamical PV (Klein and Pedlosky 1992; PT).

Briefly, if the time-averaged abyssal flow is marginally *unstable* ($\Upsilon_0 = 1.0$), an oscillatory component in the abyssal flow [e.g., $\Upsilon(T) = \mathcal{H}\cos(\omega T)$] *will not,* generically in the inviscid limit, *stabilize* the flow. Notwithstanding this generic response, however, there is a "small" set of amplitudes and frequencies (i.e., $\mathcal{H}$ and ω values) for the oscillatory component that will stabilize, even in the linear inviscid stability problem, a time-averaged marginally unstable abyssal flow. Pedlosky and Thomson (2003) have shown that there is a rather wide range of time evolutions possible for $A(T)$ as a function of the dissipation, amplitude, and frequency. The solutions for $A(T)$ are all bounded and oscillate. While the solutions can be periodic [appearing to be "monochromatic" in the way that the elliptic function solutions to the corresponding $\Upsilon(T) = \mu = 0$ are; see Fig. 4 in PT and Fig. 3 in Mooney and Swaters (1996)], they can also exhibit more complex periodic behavior with intermittent higher-frequency oscillations (see Fig. 5 in PT) and the solutions can also exhibit highly irregular oscillatory aperiodic behavior (see Figs. 6 and 7 in PT).

On the other hand, if the time-averaged abyssal flow is marginally *stable* ($\Upsilon_0 = -1.0$), an oscillatory component in the abyssal flow *does,* generically in the inviscid limit, *destabilize* the flow in the linear stability problem

(see PT). As pointed out by PT, this is an important new result that suggests that the transition to instability can occur over a much larger, and more importantly over a more realistic, range of flow parameters. Regardless, nonlinearity always acts, ultimately, to bound the evolution of $A(T)$ (with or without dissipation present).

4. Weakly nonlinear evolution of a $K = 1$ wave packet

The phase velocity of the marginally unstable $K = 1$ and $\gamma_c = 0$ mode will be given by $c = 1$ [see (3.1)]. In addition, it follows from (2.2) that the leading-order Eulerian velocity field in the abyssal layer is given by $\mathbf{u}_2 \simeq \hat{\mathbf{e}}_1$ if $\gamma_c = 0$. The steady velocity in the abyssal layer given by $\mathbf{u}_2 \simeq \hat{\mathbf{e}}_1$, which has been previously referred (Swaters 1991) to as the *Nof velocity* (Nof 1983), arises from the geostrophic adjustment of a density-driven abyssal flow lying directly on a sloping bottom. The phase velocity of the marginally unstable $K = 1$ mode is therefore identical everywhere in the abyssal layer to the Nof velocity and the entire abyssal layer forms a critical layer. As is well known (see, e.g., Benney and Bergeron 1969 or Warn and Gauthier 1989), there will be a rapid development of the dimensionality of the underlying phase space as more and more modes are excited by the fundamental harmonic because of the intrinsic nonlinearity of the critical layer.

Following Mooney and Swaters (1996), in order to examine the nonlinear evolution of the marginally unstable $K = 1$ mode, it is convenient to move into a comoving reference frame in which the frequency, to leading order in the abyssal layer, will be zero. To this end, the correct scalings for the slope of the abyssal height will be given by

$$\gamma = \varepsilon^2[\Upsilon_0 + \Upsilon(T)], 0 < \varepsilon \ll 1, \Upsilon(T) \simeq O(1), \Upsilon_0 = \pm 1,$$

$$(4.1)$$

and the perturbation fields for the marginally stable or unstable $K = 1$ and $\gamma_c = 0$ modes will scale according to

$$(\eta, h) = (\varepsilon\tilde{\eta}, \varepsilon^2\tilde{h})(\tilde{x}, y, X, T; \varepsilon), \tilde{x} = x - t, X = \varepsilon x, T = \varepsilon t.$$

$$(4.2)$$

Substituting (4.2) into the nonlinear perturbation equations in (2.4) with $r = \varepsilon\mu$ yields, after dropping the tildes and a little algebra,

$$(\Delta + 1)\eta_x = \varepsilon[(\partial_T + \mu)\Delta\eta - 2\eta_{Xxx} - \eta_X - h_x$$
$$+ J(\eta, \Delta\eta)] + O(\varepsilon^2) \quad \text{and} \quad (4.3)$$

$$(\partial_T + \partial_X + \mu)h - [\Upsilon_0 + \Upsilon(T)]\eta_x + J(\eta, h) + O(\varepsilon) = 0. \quad (4.4)$$

Note that the abyssal-layer equation in (4.4) is nonlinear to leading order.

Following Mooney and Swaters (1996), (4.3) and (4.4) can be solved with an asymptotic expansion of the form

$$(\eta, h)(x, y, X, T; \varepsilon) = (\eta_0, h_0)(x, y, X, T)$$
$$+ \varepsilon(\eta_1, h_1)(x, y, X, T) + \cdots. \quad (4.5)$$

Substitution of (4.5) into (4.3) and (4.4) leads to [after solving the $O(1)$ equations and removing secular producing terms in the $O(\varepsilon)$ equations]

$$\eta_0 = A(X, T)\sin(ly)\exp(ikx) + \text{c.c.}, \quad (4.6)$$

$$\int_0^L \int_0^{2\pi/k} \{[(\partial_T + \mu)\Delta - 2\partial_{Xxx} - \partial_X]\eta_0$$
$$- h_{0x}\}\sin(ly)\exp(-ikx)\,dx\,dy + \text{c.c.} = 0, \quad \text{and} \quad (4.7)$$

$$(\partial_T + \partial_X + \mu)h_0 - [\Upsilon_0 + \Upsilon(T)]\eta_{0x} + J(\eta_0, h_0) = 0. \quad (4.8)$$

Equations (4.6), (4.7), and (4.8) form a closed system of equations for $h_0(x, y, X, T)$ and $A(X, T)$. Writing the coupled equations in this way makes explicit the similarity of this derivation with the Warn and Gauthier (1989) analysis for marginally unstable baroclinic zonal flow in the Phillips model. If the ∂_X derivatives in (4.7) and (4.8) are neglected and it is assumed that $\Upsilon(T) = \mu = 0$, it is possible to obtain a closed-form solution in terms of elliptic and trigonometric functions (Warn and Gauthier 1989).

It appears not possible to generalize the Warn and Gauthier technique if the X dependence is retained or if $\Upsilon(T) \neq 0$. It is important to emphasize, however, that it remains an interesting and challenging problem to modify the Warn and Gauthier technique if slow space variations are retained in the wave amplitude, or $\Upsilon(T) \neq 0$, or if dissipation is retained, not only in the abyssal problem examined here but also, more generally, in the Phillips model.

A spectral solution (see Mooney and Swaters 1996; Ha 2004) for h_0 can be constructed in the form

$$h_0(X, T) = -\frac{l}{2}\sum_{n=1}^{\infty} \alpha_n(X, T)n\sin(nly)$$
$$+ \left\{\sum_{m=1}^{\infty}\sum_{n=1}^{\infty}[-i + (i+1)\delta_{m1,n1}]B_{m,n}(X, T)\right.$$
$$\left. \times \sin(nly)\exp(imkx) + \text{c.c.}\right\}. \quad (4.9)$$

If (4.9) is substituted into (4.7) and (4.8), an infinite coupled set of amplitude equations is obtained that appears to be intractable. Mooney and Swaters (1996) have described the qualitative behavior of the solutions with the dependence on X neglected and with $\Upsilon(T) = \mu = 0$ for various truncations. In general, the solutions are bounded and periodic but the qualitative structure of the periodicity becomes more complex as the number of terms in the truncation is increased.

However, in the simplest truncation in which only the fundamental mode $A(X, T)$ and the mean flow it generates $\alpha_2(X, T)$ are retained, and $B_{1,3}$ and all higher-order terms are neglected [see, e.g., Pedlosky (1972) for the Phillips model analog], it is known that the resulting model equations can be transformed into the sine–Gordon (SG) equation (Gibbon et al. 1979). The SG equation is a completely integrable solitary wave equation (Ablowitz and Segur 1981) that, in the present context, has an abyssal soliton solution. It is of interest to determine the propagation characteristics of the abyssal soliton when dissipation is present and the background marginal flow is time varying.

a. Solution of the truncated soliton model

If $B_{1,3}$ and all higher-order terms are neglected, the truncated spectral equations are given by

$$(\partial_T + \partial_X + \mu)[\partial_T + (1 - 2k^2)\partial_X + \mu]A$$
$$= k^2[\Upsilon_0 + \Upsilon(T)]A - l^2k^2AB \quad \text{and} \quad (4.10)$$

$$(\partial_T + \partial_X + \mu)B = [\partial_T + (1 - 2k^2)\partial_X + 2\mu]|A|^2, \quad (4.11)$$

where, for convenience, $B \equiv \alpha_2(X, T)$. These have been referred to as the AB equations (see, e.g., Tan and Boyd 2002).

In the absence of spatial variations (i.e., $\partial_X \equiv 0$), (4.10) and (4.11) are identical to (3.9) and (3.10), with, of course, $K = 1$. Thus, in this limit, (4.10) and (4.11) are identical to PT's (3.13a,b) where a complete discussion of the evolution of A and B when time variability and dissipation are given. Here, the effect of $\Upsilon(T) \neq 0$ and $\mu \neq 0$ on the soliton solution to (4.10) and (4.11) is determined.

The method of analysis is based on a nonlinear WKB procedure for solitary waves that assumes that the time

scale associated with $\Upsilon(T)$ and the dissipation is long in comparison with the advective time scale of the soliton. It is assumed that $\Upsilon_0 + \Upsilon(T) \equiv \tilde{\gamma}(\delta T)/k^2$ and $\mu = \nu\delta$ with $\tilde{\gamma}(\delta T) \simeq O(1)$ and $\upsilon \simeq O(1)$, where $0 < \delta \ll 1$, and that A and B are real valued and satisfy the far-field conditions $|A, B| \to 0$ as $X \to \pm\infty$ for all $T \geq 0$. The solution to (4.10) and (4.11) is constructed in the form

$$(A, B) = (A, B)(\xi, \tau; \delta), \xi = X - \frac{1}{\delta} \int_0^{\delta T} c(\zeta) \, d\zeta, \tau = \delta T,$$
(4.12)

where $c(\tau)$ is the soliton velocity.

Substitution of (4.12) into (4.10) and (4.11) leads to, after a little algebra,

$$(1 - c)(1 - c - 2k^2)A_{\xi\xi} - \gamma(\tau)A + k^2 l^2 AB$$
$$= -\delta(1 - c - 2k^2)A_{\xi\tau} - \delta[(1 - c)A_\xi]_\tau$$
$$- 2\delta\nu(1 - c - k^2)A_\xi + O(\delta^2) \quad \text{and} \quad (4.13)$$

$$(1 - c)B_\xi - (1 - c - 2k^2)(A^2)_\xi$$
$$= \delta[(A^2 - B)_\tau + \nu(2A^2 - B)], \quad (4.14)$$

where the tilde has been dropped on γ. The solution, in the limit of "small" δ, can be found in the form

$$(A, B) = (A, B)^{(0)} + \delta(A, B)^{(1)} + \cdots. \quad (4.15)$$

Substitution of (4.15) into (4.13) and (4.14) leads to the following series of problems.

1) THE $O(1)$ PROBLEM

The $O(1)$ equations can be written in the following form, after a little algebra:

$$(1 - c)(1 - c - 2k^2)A_{\xi\xi}^{(0)} - \gamma A^{(0)}$$
$$+ \frac{k^2 l^2 (1 - c - 2k^2)[A^{(0)}]^3}{(1 - c)} = 0 \quad \text{and}$$
(4.16)

$$B^{(0)} = \frac{(1 - c - 2k^2)[A^{(0)}]^2}{(1 - c)}. \quad (4.17)$$

It is straightforward to verify that (4.16) has the soliton solution

$$A^{(0)}(\xi, \tau) = A_0(\tau) \operatorname{sech}[\kappa(\tau)\xi], \quad (4.18)$$

where

$$\kappa(\tau) \equiv \sqrt{\gamma(\tau)/\{[1 - c(\tau)][1 - c(\tau) - 2k^2]\}} \quad \text{and}$$
$$A_0(\tau) \equiv \sqrt{2\gamma(\tau)[1 - c(\tau)]/\{k^2 l^2 [1 - c(\tau) - 2k^2]\}},$$
(4.19)

which implies that A_0 and κ are related through the simple algebraic relation $A_0 = \sqrt{2} (1 - c) \kappa/(kl)$. Thus, given $\gamma(\tau)$ and $c(\tau)$, the evolution of the soliton "wavenumber" $\kappa(\tau)$ and "amplitude" $A_0(\tau)$ will be determined. The parameter $\gamma(\tau)$ is assumed known and $c(\tau)$ remains to be determined.

A slowly varying phase shift parameter can be introduced into (4.18). However, the evolution of the phase shift is determined by higher-order solvability conditions (Kodama and Ablowitz 1980) that are beyond the scope of this paper. Consequently, for the present purposes, any phase shift can be assumed constant and "absorbed" into ξ.

For a solution of the form (4.18) to be bounded, it is required that the product,

$$(1 - c) \times (1 - c - 2k^2),$$

must have the same sign as γ or else κ and A_0 will be imaginary and (4.18) will be unbounded whenever $\kappa\xi = (2n + 1) i\pi/2$ for $n \in \mathbb{Z}$. Thus, for the marginally unstable ($\gamma > 0$) situation,

$$c \in (-\infty, 1 - 2k^2) \cup (1, \infty),$$

and for the marginally stable ($\gamma < 0$) situation,

$$c \in (1 - 2k^2, 1).$$

The allowed sets of translation velocities associated with the marginally unstable and stable situations are disjoint from one another (see, also, Pedlosky 1972; Gibbon et al. 1979).

The boundaries between the unstable and stable regions, given by $c = 1 - 2k^2$ and $c = 1$, have an important role to play in terms of the path in parameter space that a perturbed soliton can evolve along. It follows from (4.18) and (4.19) that $A^{(0)} \to 0$ for all ξ (i.e., for all X) as $c \to 1$. Moreover, it can be shown that the limit of each of the three terms in (4.16) is zero as $c \to 1$ for the soliton solution (4.18). Thus the limit

$$\lim_{c \to 1} A^{(0)}(\xi, \tau) = 0,$$

is itself a solution of (4.16) if $\gamma \neq 0$. It will be shown below that the $c = 1$ soliton is the end state toward which all slowly dissipating solitons evolve.

However, the limit of $A^{(0)}(\xi, \tau)$ as $c \to 1 - 2k^2$ is *not* a (classical) solution to (4.16). It follows from (4.18) and (4.19) that $A^{(0)} \to 0$ except for when $\xi = 0$ where $A^{(0)} \to \infty$ (the limits are not interchangeable) as $c \to 1 - 2k^2$ (assuming $\gamma \neq 0$). This means that the limit

$$\lim_{c \to 1 - 2k^2} A^{(0)}(\xi, \tau),$$

does not correspond to, and cannot be achieved by, a "slowly varying" soliton. The value $c = 1 - 2k^2$ forms a "barrier" in parameter space across which a deforming soliton cannot directly cross. The consequence is that, for example, a dissipating soliton with $c(0) < 1 - 2k^2$ cannot evolve directly toward the $c = 1$ zero solution (because it cannot pass through $c = 1 - 2k^2$) and a more complicated path in parameter space must be followed. This is shown below.

2) THE DETERMINATION OF $c(\tau)$

The evolution of $c(\tau)$ can be determined by the application of solvability conditions associated with the $O(\delta)$ equations. Specifically, the required solvability condition is that the inhomogeneous terms must be orthogonal to the kernel (i.e., the vector space spanned by the homogeneous solutions) of the corresponding adjoint system (see, e.g., Kodama and Ablowitz 1980; Swaters and Flierl 1991) associated with the $O(\delta)$ problem. The result of this solvability condition is to derive an ordinary differential equation for $c(\tau)$. An alternate, and more physically appealing but a completely equivalent mathematical, viewpoint is to derive the governing equation for $c(\tau)$ based on phased-averaged conservation laws (see, e.g., Grimshaw 1979a,b). This is the approach taken here.

The energy equation associated with (4.10) and (4.11) can be derived from $2(A_T + A_X) \times$ (4.10), which can be written in the form

$$[(A_T + A_X)^2 - \gamma A^2 + l^2 k^2 A^2 (B - A^2/2)]_T$$
$$+ \{(1 - 2k^2)(A_T + A_X)^2 - \gamma A^2$$
$$+ l^2 k^2 A^2 [B - (1 - 2k^2)A^2/2]\}_X$$
$$= -\gamma_T A^2 + \nu\delta\{l^2 k^2 A^2 (2A^2 - B)$$
$$- 4(A_T + A_X)[A_T + (1 - k^2)A_X]\} + O(\nu^2).$$
(4.20)

If (4.12) and (4.15) are substituted into (4.20), it follows that to leading order,

$$\frac{\partial}{\partial \tau}\left\{ \int_{-\infty}^{\infty} (1 - c)^2 [A_\xi^{(0)}]^2 - \gamma [A^{(0)}]^2 \right.$$
$$\left. + \frac{l^2 k^2 (1 - c - 4k^2)}{2(1 - c)} [A^{(0)}]^4 \, d\xi \right\}$$
$$= \int_{-\infty}^{\infty} \left\{ \frac{\nu l^2 k^2 (1 - c + 2k^2)}{(1 - c)} [A^{(0)}]^4 - \gamma_\tau [A^{(0)}]^2 \right.$$
$$\left. - 4\nu(1 - c)(1 - c - k^2)[A_\xi^{(0)}]^2 \right\} d\xi, \quad (4.21)$$

which, if (4.17), (4.18), and (4.19) are exploited, simplifies to, after some algebra,

$$\frac{d}{d\tau}\left[\frac{\gamma(1 - c)}{(1 - c - 2k^2)}\right] = -\frac{4\nu\gamma(1 - c)}{(1 - c - 2k^2)} + \gamma_\tau(1 - c)/k^2.$$
(4.22)

b. Description of $c(\tau)$

It is possible to obtain an analytical solution, within quadrature, for $c(\tau)$ from (4.22) if $\gamma_\tau \neq 0$ (i.e., the marginally unstable or stable flow has time variability) *and* $\nu \neq 0$ (i.e., dissipation is present). However, the highly implicit form of the solution does not easily lend itself to qualitative description. It is more useful to consider the two sublimits $\gamma_\tau = 0$ and $\nu = 0$ individually and describe those solutions. Because (4.10) and (4.11) (without dissipation and time variability in γ) is a *soliton* model and solvable with the inverse scattering transform (Ablowitz and Segur 1981), thusly depending continuously on the initial data, the evolution of the soliton with both *weak* dissipation and time variability present will depend smoothly on γ_τ and ν from one sublimit to the other (Kaup and Newell 1978). There will be no combination of parameter values for which the qualitative evolution of $c(\tau)$ diverges significantly from either of these two limits.

1) THE DISSIPATIVE SOLUTION WITHOUT TIME-VARYING MARGINAL FLOW

In the limit $\gamma_\tau = 0$ [i.e., $\Upsilon(T) = 0$ so that $\gamma = k^2$ or $\gamma = -k^2$ in the marginally unstable or stable situations, respectively], (4.22) can be solved to give

$$\frac{1 - c}{1 - c - 2k^2} = \left(\frac{1 - c_0}{1 - c_0 - 2k^2}\right) \exp(-4\nu t),$$
(4.23)

or, equivalently,

$$c(\tau) = 1 + \frac{2k^2(1 - c_0)/(1 - c_0 - 2k^2)}{\exp(4\nu t) - (1 - c_0)/(1 - c_0 - 2k^2)},$$
(4.24)

where $c_0 = c(0)$. It follows from (4.23) that the soliton amplitude $A_0(\tau)$, as determined by (4.19b), will decay exponentially in time irrespective of the value of c_0 (assuming $c_0 \neq 1$ or $c_0 \neq 1 - 2k^2$). If $c_0 = 1$, then $A^{(0)}(\xi, \tau) = 0$ for all $\tau \geq 0$. As described previously, there is no smooth solution if $c_0 = 1 - 2k^2$, so this possibility is not relevant.

However, because of the singularity associated with $c_0 = 1 - 2k^2$ in (4.23), the details of the evolution of the

translation velocity $c(\tau)$ and the wavenumber $\kappa(\tau)$ do depend on c_0. It is easy to verify that

$$\frac{1 - c_0}{1 - c_0 - 2k^2} < 1 \Leftrightarrow c_0 > 1 - 2k^2. \quad (4.25)$$

It therefore follows from (4.24) that if $c_0 > 1 - 2k^2$, then $c(\tau) \to 1$ monotonically as τ increases. Note that in this case there is no value of τ for which the denominator in (4.24) is zero (it will always be positive). In addition, in this case, it follows from (4.19a) that $\kappa(\tau) \to \infty$ monotonically as $\tau \to \infty$. Since $A_0(\tau) \to 0$ and $\kappa(\tau) \to \infty$ monotonically, it follows from (4.18) that $A^{(0)}(\xi, \tau) \to 0$, as $\tau \to \infty$. Indeed, since $\kappa(\tau) \to \infty$, the ξ region over which $A^{(0)}(\xi, \tau)$ is significantly different than zero [i.e., the "support" of $A^{(0)}$] monotonically decreases over time. That is, in this case, the soliton amplitude and horizontal extent monotonically decreases over time, and its translation velocity approaches $+1.0$, at which point the soliton has dissipated to zero.

However, if $c_0 < 1 - 2k^2$ (which can only occur in the marginally unstable $\gamma > 0$ case), the evolution in parameter space for the dissipating soliton is more complex. This is because the soliton translation velocity cannot cross the point $c = 1 - 2k^2$. Nevertheless, the final state of the decaying soliton is the $c = 1$ zero solution.

It follows from (4.24) and (4.25) that if $c_0 < 1 - 2k^2$, $c(\tau)$ monotonically *decreases* and, in fact,

$$c(\tau) \to -\infty \text{ as } \tau \uparrow \tau_* \equiv \frac{1}{4\nu} \ln[(1 - c_0)/(1 - c_0 - 2k^2)] > 0.$$

Concomitantly, it follows from (4.19) that

$$\lim_{\tau \to \tau_*} \kappa(\tau) = 0 \quad \text{and} \quad \lim_{\tau \to \tau_*} A_0(\tau) = \sqrt{2\gamma}/|kl|,$$

so that, formally, it is possible to define $A_0(\tau_*) = \sqrt{2\gamma}/|kl|$ and $\kappa(\tau_*) = 0$. Consequently, there will be a well-defined solution for $0 < \tau \leq \tau_*$. Thus, in the interval $0 < \tau \leq \tau_*$ for the marginally unstable $c_0 < 1 - 2k^2$ case, the soliton deaccelerates, its amplitude exponentially decays to a finite nonzero value, and it horizontally "spreads out" or spatially dilates with respect to X.

Moreover, in finite time, the translation velocity becomes, formally, negatively unbounded and the soliton has dilated to the point where there is no slope anywhere [i.e., $|A_X^{(0)}| \to 0$ for all X as $\tau \uparrow \tau_*$]. Indeed, the fact that the speed of the soliton is infinite at precisely the moment that $A_X^{(0)} = 0$ everywhere means there is no physically relevant signal propagation associated with this singularity at that moment.

Notwithstanding the apparent peculiarity of the result for the velocity associated with $\tau \uparrow \tau_*$, the solution exists at $\tau = \tau_*$ and continues to evolve for $\tau > \tau_*$. For $\tau > \tau_*$, (4.24) implies that $c(\tau)$ continues to decrease monotonically (and is finite the moment $\tau > \tau_*$), although now $c(\tau) > 1$ since the formal right-hand limit is $c(\tau) \to +\infty$ as $\tau \downarrow \tau_*$. In addition, since $|A_X^{(0)}| > 0$ for $\tau > \tau_*$, the soliton "*sech*" shape reemerges. In fact, $c(\tau) \to 1$, $\kappa(\tau) \to \infty$, and $A_0(\tau) \to 0$ as $\tau_* < \tau \to \infty$. That is, the moment $\tau > \tau_*$ and subsequently increases, the soliton slows down, the amplitude continues to decay exponentially and its horizontal extent contracts, and, eventually, the zero amplitude state is reached. Indeed, the scenario for $\tau > \tau_*$ is identical to the $c_0 > 1 - 2k^2$ decay description given above.

Figures 3a–c depict the evolution of the soliton velocity $c(\tau)$, amplitude $A_0(\tau)$, and wavenumber $\kappa(\tau)$, respectively, when dissipation is present. To be specific, $\nu = 1.0$ and $k = \sqrt{3}/2$ (thus, $l = 1/2$ and $1 - 2k^2 = -1/2$) are chosen. The three graphs in each figure correspond to the initial conditions $c_0 = 2.0$, 0.5, and -1.0, respectively. The $c_0 = 2.0$ and $c_0 = -1.0$ initial conditions are associated with the marginally *unstable* flow $\gamma = 1/k^2 = 4/3$ with $c_0 = 2.0 > 1 - 2k^2$ and $c_0 = -1.0 < 1 - 2k^2$. The $c_0 = 0.5$ initial condition is associated with the marginally *stable* flow $\gamma = -1/k^2 = -4/3$ (which must always satisfy $1 - 2k^2 < c_0 < 1$).

Figure 3a shows the monotonic decay of $c(\tau)$ toward 1.0 as τ increases associated with the marginally stable initial condition $c_0 = 0.5$ and the marginally unstable initial condition $c_0 = 2.0 > 1 - 2k^2$, respectively. The discontinuous evolution of $c(\tau)$ associated with marginally unstable initial condition $c_0 = -1.0 < 1 - 2k^2$ is seen. For $c_0 = -1.0$, $\nu = 1.0$, and $k = \sqrt{3}/2$, it follows that $\tau_* = \ln(2)/2$.

Figure 3b shows the continuous exponential decay in $A_0(\tau)$ associated with all three initial conditions. Figure 3c shows the exponential growth of $\kappa(\tau)$ associated with the marginally stable initial condition $c_0 = 0.5$ and the marginally unstable initial condition $c_0 = 2.0 > 1 - 2k^2$, respectively. The decay of $\kappa(\tau)$, associated with the marginally unstable initial condition $c_0 = -1.0 < 1 - 2k^2$, in the interval $0 < \tau < \tau_*$ is seen with $\kappa(\tau_*) = 0$. Thereafter $\kappa(\tau)$ increases exponentially rapidly.

2) THE INVISCID SOLUTION WITH TIME-VARYING MARGINAL FLOW

In the inviscid limit $\nu = 0$, (4.22) can be solved to give

$$\frac{(1 - c - 2k^2)^3}{\gamma^3(1 - c)(1 - c - 3k^2)^2} = \frac{(1 - c_0 - 2k^2)^3}{\gamma_0^3(1 - c_0)(1 - c_0 - 3k^2)^2}$$

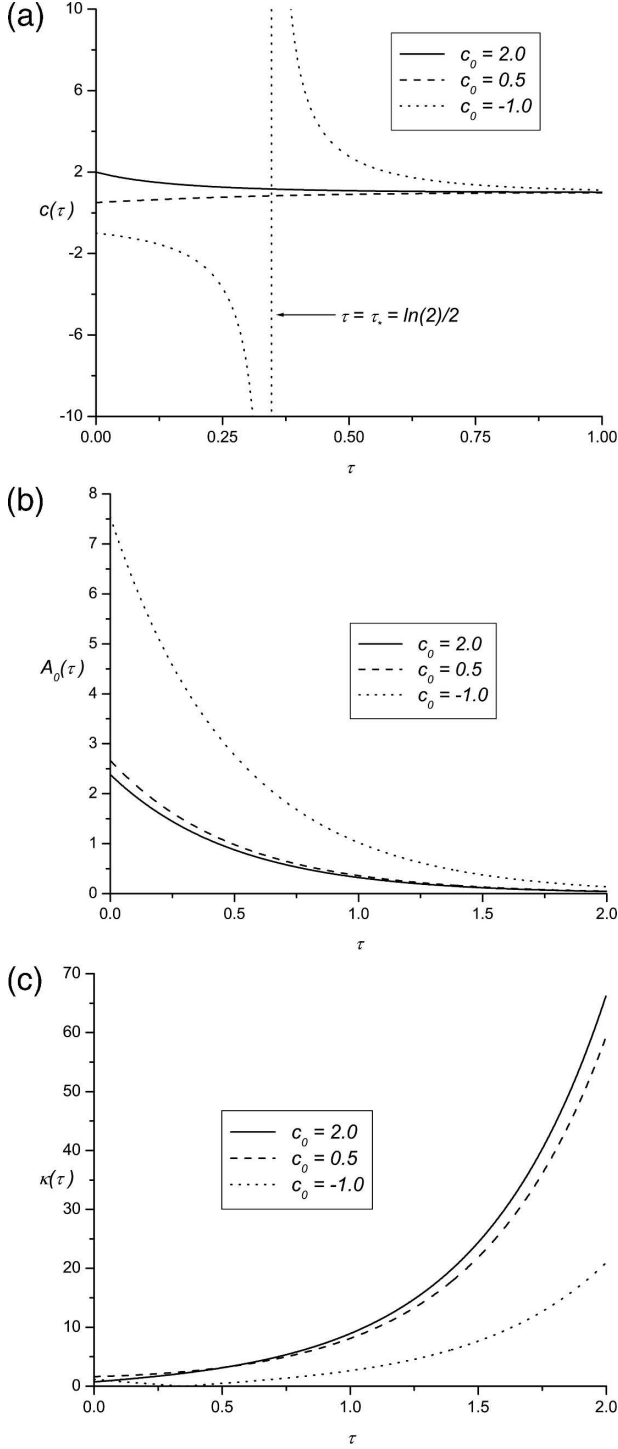$$\equiv \alpha > 0, \quad (4.26)$$

FIG. 3. The evolution of the soliton parameters [(a) velocity c, (b) amplitude A_0, and (c) wavenumber κ] with dissipation present where $\nu = 1.0$ and $k = \sqrt{3}/2$. The initial conditions $c_0 = -1 \in (-\infty, 1 - 2k^2)$ and $2 \in (1 - 2k^2, \infty)$ correspond to the marginally unstable mode with $\gamma = 1/k^2$, and $c_0 = 0.5 \in (1 - 2k^2, 1)$ corresponds to the marginally stable mode with $\gamma = -1/k^2$.

which can be written in the form

$$(1 - \alpha\gamma^3)(c - 1 + 2k^2)^3 + 3\alpha\gamma^3 k^4(c - 1 + 2k^2)$$

$$+ 2\alpha\gamma^3 k^6 = 0, \tag{4.27}$$

where $\gamma_0 = \gamma(0)$. The discriminate associated with the cubic (4.27) is given by

$$\mathcal{D} = k^{12}\alpha^2\gamma^6/(1 - \alpha\gamma^3)^3.$$

For the marginally stable case in which $\gamma < 0$, it follows that $\mathcal{D} > 0$ so that there will exist only one real solution to (4.27) for $c(\tau)$, which is the physical root. Moreover, this solution will be continuous and exist for all $\tau \geq 0$. As one simple example, suppose that $k = \sqrt{3}/2$ and that

$$\gamma(\tau) = -1 - \sin(\tau)/2 < 0 \quad \text{and} \quad c_0 = 0.5 \in (1 - 2k^2, 1),$$

which implies that $\alpha \simeq 0.65$, $A_0(0) \simeq 2.31$, and $\kappa(0) \simeq 1.41$. Figures 4a–c show the periodic evolution of the soliton parameters $c(\tau)$, $A_0(\tau)$, and $\kappa(\tau)$, respectively. The solutions are periodic with period very close, but not exactly equal, to 2π. In addition, the solutions are not purely sinusoidal [i.e., they are not solely proportional to $\sin(\tau)$ and/or $\cos(\tau)$]. The solutions are not monochromatic because, to begin with, $c(\tau)$, as a solution to the cubic (4.27), does not depend precisely linearly on $\gamma(\tau)$, and $A_0(\tau)$ and $\kappa(\tau)$ do not depend linearly on $c(\tau)$.

There are any number of time dependencies that can be examined for $\gamma(\tau) < 0$ in addition to the periodic case just described. One interesting situation corresponds to a marginally stable abyssal current with negative velocity that is accelerating toward zero velocity. That is, to describe the case where $\gamma \uparrow 0$ as τ increases, assuming $\gamma_0 < 0$ (the speed of the flow is decreasing). The marginally unstable case where $\gamma_0 > 0$ is described below and is quite different.

It follows from (4.27) that $\gamma = 0 \Rightarrow c = 1 - 2k^2$. This is, however, a singular limit in the sense that there is no uniquely defined solution to (4.16) for these parameter values, since the coefficients associated with each individual term will be identically zero. Thus, any (twice-continuously differentiable) function is a solution to (4.16) in the limit $\gamma \to 0$ and $c \to 1 - 2k^2$.

Nevertheless, it can be shown, using (4.27), that

$$c \simeq 1 - 2k^2 - \gamma k^2 \sqrt[3]{2\alpha} + O(\gamma^3) \quad \text{as} \quad \gamma \to 0,$$

which implies, using (4.19), that

$$\lim_{\gamma \to 0} \kappa = \tilde{\kappa} = 2^{-2/3} k^{-2} \alpha^{-1/6} \quad \text{and}$$

$$\lim_{\gamma \to 0} A_0 = \tilde{A}_0 = 2^{5/6} \alpha^{-1/6}/|kl|, \tag{4.28}$$
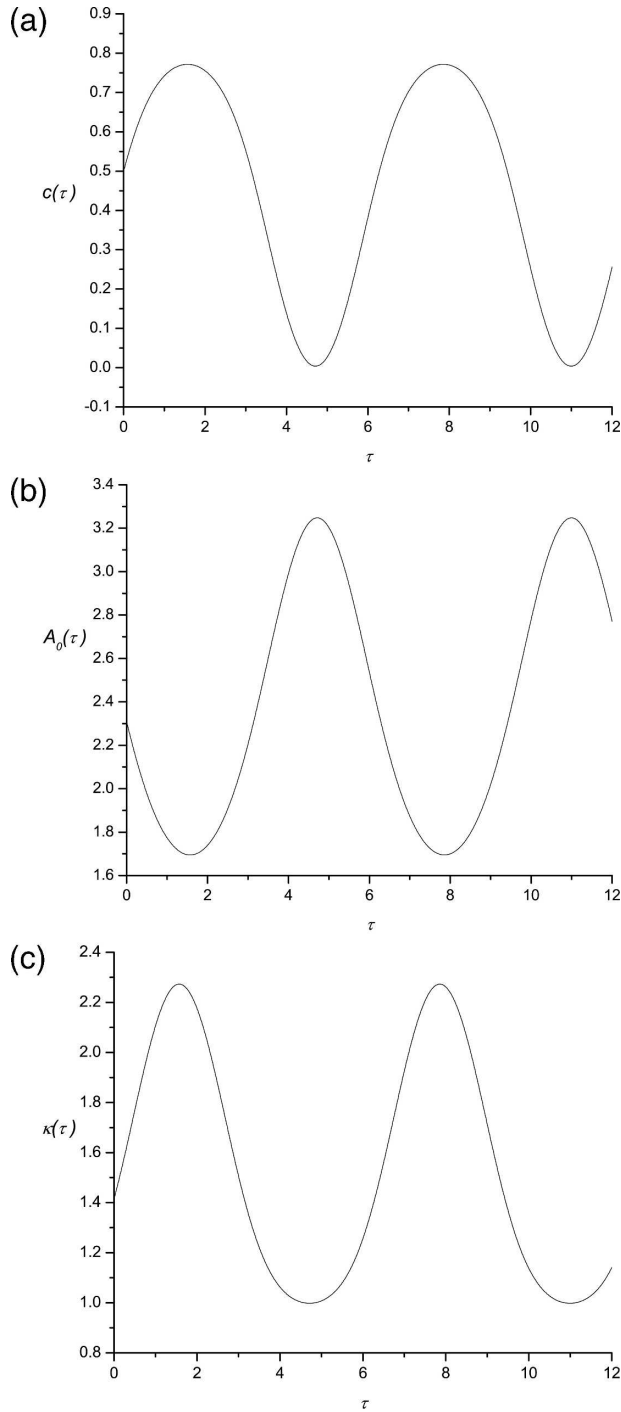
(a)

(b)

(c)

FIG. 4. The evolution of the soliton parameters [(a) velocity c, (b) amplitude A_0, and (c) wavenumber κ] assuming marginally stable periodic inviscid flow given by $\gamma = -1 - \sin(\tau)/2$. The initial soliton velocity is $c_0 = 0.5$ and $k = \sqrt{3}/2$, which implies $A_0(0) \simeq 2.31$ and $\kappa(0) \simeq 1.41$. The evolution is periodic with a period of about 2π.
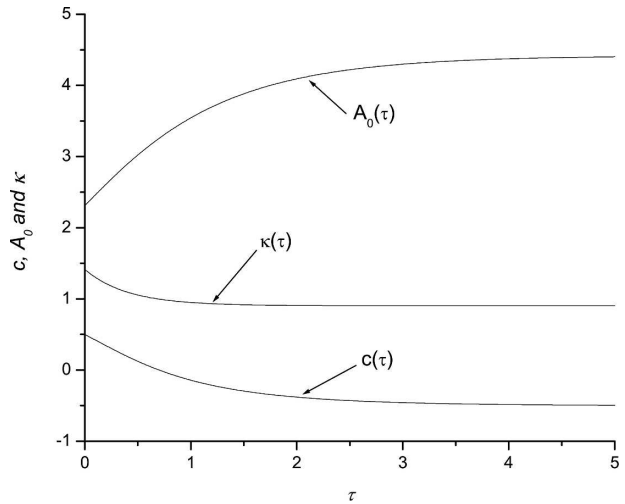

FIG. 5. The evolution of the soliton parameters c, A_0, and κ assuming a marginally stable flow that is accelerating toward zero velocity given by $\gamma = -\exp(-\tau)$. The initial soliton parameters are the same as in Fig. 4. The parameters evolve to a constant value implying that even though the background flow velocity is zero, a soliton continues to steadily propagate.

which in turn implies that

$$\lim_{\gamma \to 0} A^{(0)}(\xi, \tau) = 2^{5/6} \alpha^{-1/6} \operatorname{sech}(2^{-2/3} k^{-2} \alpha^{-1/6} \xi)/|kl|.$$

(4.29)

This is an important result because it suggests that should a wave packet soliton be formed on a marginally stable abyssal current, with the mean flow speed decreasing to zero, the solution does *not* tend to zero amplitude but rather continues to propagate in the limiting form given by (4.28) and (4.29). In the situation in which $\gamma(\tau)$ passes through zero and becomes positive, the $\gamma < 0$ solution smoothly connects with a $\gamma > 0$ solution (as described below). Figure 5 shows the monotonic evolution of the soliton parameters toward this limiting solution assuming $\gamma(\tau) = -\exp(-\tau)$ with $k = \sqrt{3}/2$ and $c_0 = 0.5$. For these parameter values the limiting soliton wavenumber and amplitude are given by $\tilde{\kappa} \simeq 0.9$ and $\tilde{A}_0 \simeq 4.42$, respectively.

In the marginally unstable situation $\gamma > 0$, the solutions for $c(\tau)$, $A_0(\tau)$, and $\kappa(\tau)$ have more complex behavior than those associated with $\gamma < 0$. Mathematically, this is a consequence of the fact that the discriminate $\mathcal{D}$ is of indeterminate sign for $\gamma > 0$ and is undefined when $\gamma^3 = 1/\alpha$. However, in the situations in which $0 < \alpha\gamma^3 < 1$ *or* $\alpha\gamma^3 > 1$ for all $\tau \geq 0$, $\mathcal{D}$ is positive or negative definite, respectively, and the solutions will be qualitatively similar to those associated with $\gamma < 0$ and will not be further described. The most interesting situation occurs when it is possible that $\gamma^3 = 1/\alpha$.

To begin, it is noted that when $\gamma^3 = 1/\alpha$ for some value of τ, the only *finite* solution to (4.27) is $c = 1 - 8k^2/3$. Formally, the other two solutions of (4.27) become unbounded as $\gamma^3 \to 1/\alpha$. The principal property that determines the qualitative evolution of $c(\tau)$ is whether or not, for a given initial condition c_0, $c(\tau) \to 1 - 8k^2/3$ as $\gamma^3 \to 1/\alpha$.

As shown previously, if $\gamma > 0$, then $c \in (-\infty, 1 - 2k^2) \cup (1, \infty)$ [or else κ and A_0 are imaginary and $A^{(0)}$ is periodically unbounded]. For the subregion

$$c_0 \in I_1 \equiv (-\infty, 1 - 3k^2) \cup (1, \infty),$$

the solution $c(\tau)$ is unable to reach $1 - 8k^2/3$ since the left-hand side of (4.26) has "vertical asymptotes" at $c = 1 - 3k^2$ and $c = 1$, respectively. Moreover, for $c_0 \in I_1$, $|c(\tau)| \to \infty$ as $\gamma^3 \to 1/\alpha$ and there is no further time evolution. Consequently, for $c_0 \in I_1$, it follows from (4.18) and (4.19) that

$$\kappa \to 0 \quad \text{and} \quad A_0 \to \hat{A}_0 \equiv \sqrt{2}\alpha^{-1/6}/|kl| \text{ as } \gamma^3 \to 1/\alpha,$$
$$(4.30)$$

so that the soliton evolves to a constant solution without any spatial shear *and does so in finite time.* The only set of initial conditions that allow the *possibility* that $c(\tau) \to 1 - 8k^2/3$ as $\gamma^3 \to 1/\alpha$ are the following:

$$c_0 \in I_2 \equiv (1 - 3k^2, 1 - 2k^2).$$

To further illustrate the properties of the $\gamma > 0$ solutions it is useful to work with a specific example. It is assumed that $\gamma(\tau) = \exp(-\tau)$, which corresponds to a marginally unstable flow with positive velocity that is deaccelerating toward zero velocity (i.e., $\gamma \downarrow 0$ as $\tau \to \infty$). Since $\gamma \leq 1$, the set of initial conditions for which $\gamma^3 = 1/\alpha$ for some $\tau > 0$ corresponds to those c_0 for which $\alpha > 1$. It follows from (4.26) that $\alpha > 1$, *if and only if,*

$$c_0 \in (-\infty, 1 - 3k^2) \cup (1 - 3k^2, 1 - 8k^2/3) \cup (1, \infty).$$
$$(4.31)$$

Note that the union of the first and third subintervals in (4.31) is exactly I_1.

The second subinterval in (4.31) is a subset of I_2. For the remaining initial conditions in I_2, given by

$$c_0 \in I_4 \equiv (1 - 8k^2/3, 1 - 2k^2),$$

(which are, in fact, the only remaining c_0 allowed for $\gamma > 0$), it follows that $\alpha < 1$ and there is no value of τ for which $\gamma^3 = \exp(-3\tau) = 1/\alpha > 1$. For $c_0 \in I_4$, $\mathcal{D} > 0$ and $c(\tau)$ evolves continuously and qualitatively behaves like the $\gamma < 0$ solutions. Physically, if $c_0 \in I_4$, then only if the background abyssal flow is *accelerating,* that
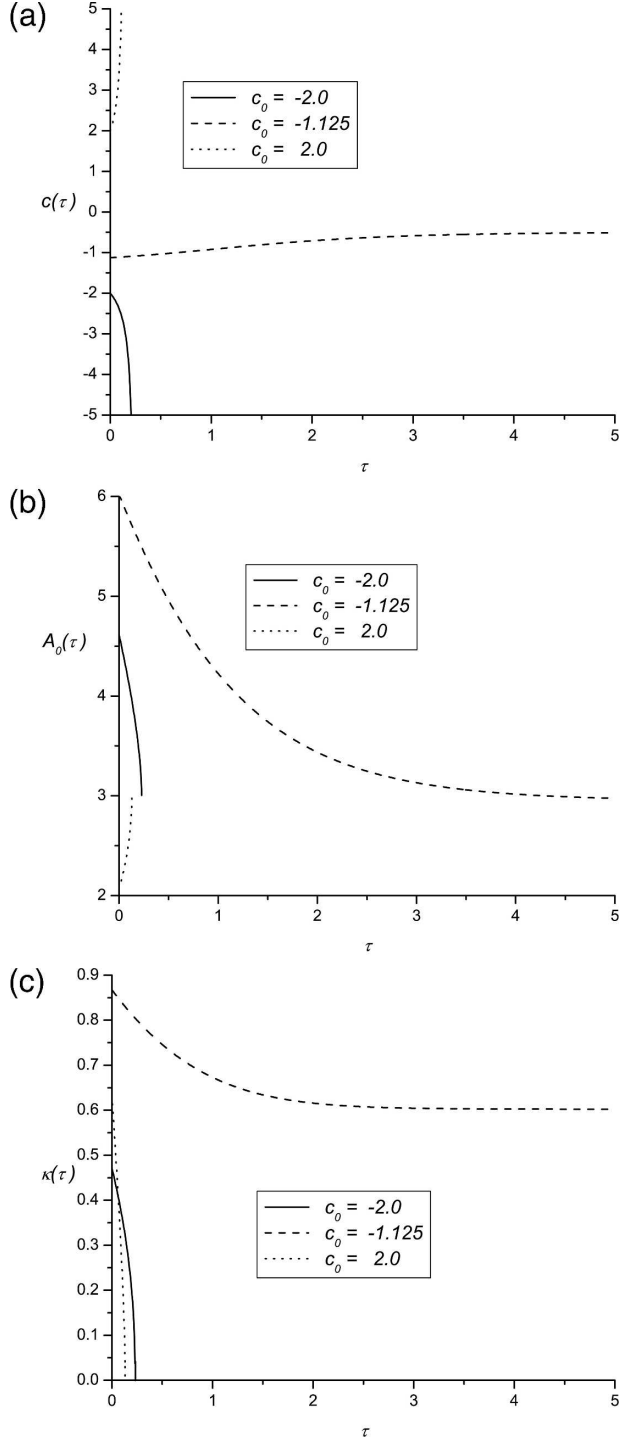

FIG. 6. The evolution of the soliton [(a) velocity c, (b) amplitude A_0, and (c) wavenumber κ] for a marginally unstable flow deaccelerating toward zero velocity with $\gamma = \exp(-\tau)$, $k = \sqrt{3}/2$, and $\nu = 0$. The initial conditions $c_0 = -2.0$, -1.125, and 2.0 correspond to the subintervals in (4.31). The $c_0 = -2.0$ and 2.0 solitons evolve to a constant solution in finite time. The $c_0 = -1.125$ soliton can connect with a marginally stable soliton.

is, $\gamma > 0$ and $\gamma_\tau > 0$, can there be values of τ for which $\gamma^3 = I/\alpha$.

For the deaccelerating flow (i.e., $\gamma > 0$ and $\gamma_\tau < 0$) examined here, for $c_0 \in I_2$ (which includes $c_0 \in I_4$), the solutions *all* evolve continuously toward the limiting solution $c = 1 - 2k^2$ with the soliton wavenumber and amplitude given by (4.28) as $\gamma \to 0$. These are the solutions that can smoothly connect with the $\gamma < 0$ solution across $\gamma = 0$.

To further illustrate the solution associated with the $\gamma = \exp(-\tau)$ example it is assumed that $k = \sqrt{3}/2$ and $c_0 = \{-2, -1.125, 2\}$. These three initial conditions are, respectively, in each of the three subintervals in (4.31). In light of the above remarks, the solutions associated with $c_0 = -2$ and 2 will have the property that $|c(\tau)| \to \infty$ and $\kappa(\tau)$ and $A_0(\tau)$ will satisfy (4.30) as

$$\tau \uparrow \tau_* \equiv \ln(\alpha)/3 > 0,$$

and there will be no further time evolution associated with these solutions for $\tau > \tau_*$. The solution associated with $c_0 = -1.125$ will smoothly evolve to $c(\tau_*) = 1 - 8k^2/3 = -1$ and will subsequently evolve to $c(\tau) = 1-2k^2 = -1/2$ as $\tau \to \infty$. It is, generically, this solution that connects with a $\gamma < 0$ solution if γ were to change sign.

For $c_0 = 2$, it follows that $\alpha \simeq 1.48$, $A_0(0) \simeq 2.07$, $\kappa(0) \simeq 0.63$, $\tau_* \simeq 0.13$, and the limiting soliton amplitude is given by $\hat{A}_0 \simeq 3.06$. For $c_0 = -2$, it follows that $\alpha = 2.0$, $A_0(0) \simeq 4.62$, $\kappa(0) \simeq 0.47$, $\tau_* \simeq 0.23$, and the limiting soliton amplitude is given by $\hat{A}_0 \simeq 2.91$. For $c_0 = -1.125$, it follows that $\alpha \simeq 7.35$, $A_0(0) \simeq 6.02$, $\kappa(0) \simeq$ 0.87, $\tau_* \simeq 0.66$ (although there is no singular behavior at τ_*), and the limiting soliton wavenumber and amplitude, as $\tau \to \infty$, are given by $\tilde{\kappa} \simeq 0.6$ and $\tilde{A}_0 \simeq 2.95$, respectively.

Figures 6a–c show the evolution of the soliton translation velocity, amplitude, and wavenumber, respectively, associated with this example. Figure 6a depicts the rapid development of $|c(\tau)| \to \infty$ as $\tau \uparrow \tau_*$ associated with $c_0 = -2$ and 2, respectively, and the smooth evolution $c(\tau) \to 1 - 2k^2 = -1/2$ as $\tau \to \infty$ associated with $c_0 = -1.125$. Figure 6b shows the development of the soliton amplitude $A_0(\tau)$. As $\tau \uparrow \tau_*$, for $c_0 = -2$ and 2, respectively, $A_0(\tau) \to \hat{A}_0$ as determined by (4.30) and listed above, and the smooth evolution $A_0(\tau) \to \tilde{A}_0$, as determined by (4.28) and listed above, as $\tau \to \infty$ associated with $c_0 = -1.125$. Figure 6c shows the development of the soliton wavenumber $\kappa(\tau)$. As $\tau \uparrow \tau_*$, for $c_0 = -2$ and 2, respectively, $\kappa(\tau) \to 0$ as determined by (4.30) and the smooth evolution $\kappa(\tau) \to \tilde{\kappa}$, as determined by (4.28) and listed above, as $\tau \to \infty$ associated with $c_0 = -1.125$. It is the smooth $c_0 = -1.125$ solutions (i.e., initial conditions $c_0 \in I_2$) that exist for all $\tau \geq 0$, as depicted in Figs. 6a–c, that will, generically, connect to, respectively, a soliton translation velocity, amplitude, and wavenumber in the $\gamma < 0$ region, if γ were to change sign.

c. Solution to the $O(\delta)$ equations

The $O(\delta)$ equations can be written in the following form, after a little algebra:

$$\left\{ (1 - c)(1 - c - 2k^2)\partial_{\xi\xi} - \gamma + \frac{3k^2l^2(1 - c - 2k^2)}{(1 - c)}[A^{(0)}]^2 \right\} A^{(1)} = -[(1 - c)A_\xi^{(0)}]_\tau$$

$$- (1 - c - 2k^2)A_{\xi\tau}^{(0)} - 2\nu(1 - c - k^2)A_\xi^{(0)} - \frac{k^2l^2A^{(0)}\Psi}{(1 - c)} \quad \text{and} \tag{4.32}$$

$$B^{(1)} = \frac{2(1 - c - 2k^2)}{(1 - c)}A^{(0)}A^{(1)} + \frac{\Psi}{(1 - c)}, \tag{4.33}$$

where

$$\Psi \equiv \int_{\text{sgn}(c)\infty}^{\xi} \langle \{[A^{(0)}(\zeta, \tau)]^2 - B^{(0)}(\zeta, \tau)\}_\tau + \nu\{2[A^{(0)}(\zeta, \tau)]^2 - B^{(0)}(\zeta, \tau)\} \rangle \, d\zeta$$

$$= \frac{2(\gamma_\tau + \nu\gamma)}{\kappa k^2 l^2}[\tanh(\kappa\xi) - \text{sgn}(c)] + 4l^{-2}(1 - c)\kappa\kappa_\tau\xi \, \text{sech}(\kappa\xi), \tag{4.34}$$

where $\text{sgn}(c)$ is the *sign* of c [i.e., $\text{sgn}(c) = +1$ or -1 for $c > 0$ or $c < 0$, respectively; this ensures that $\Psi \to 0$ as $\xi \to \text{sgn}(c) \infty$ *ahead* of the propagating soliton] and (4.17), (4.18), (4.19), and (4.22) have been used. Writ-

ten in this form, $A^{(1)}$ is obtained from (4.32) and $B^{(1)}$ is determined from the algebraic (4.33).

In comparing with (4.16), a homogeneous solution of the self-adjoint Eq. (4.32) is given by $A_\xi^{(0)}$. The phase-averaged energy balance (4.21) can be obtained by multiplying (4.32) by $2(1 - c)A_\xi^{(0)}$ and integrating with respect to $\xi \in (-\infty, \infty)$ exploiting, where necessary, (4.16) (it is easier to work with the integral representation for Ψ).

Note that

$$\lim_{\xi \to -\,\mathrm{sgn}(c)\infty} \Psi = -4\,\mathrm{sgn}(c)(\gamma_\tau + \nu\gamma)/(\kappa k^2 l^2).$$
(4.35)

This means that the expansion (4.12) and (4.15) is not uniformly valid with respect to X. In the jargon of the perturbation theory for solitary waves, a "shelf region" has emerged behind the propagating soliton (see Knickerbocker and Newell 1980; Kodama and Ablowitz 1980). The shelf region develops because the slowly varying soliton is unable to simultaneously satisfy the leading-order energy balance relation (4.21) *and* the "mass" balance relations,

$$\int_{-\infty}^{\infty} k^2 l^2 A^{(0)}(\xi, \tau)B^{(0)}(\xi, \tau) - \gamma(\tau)A^{(0)}(\xi, \tau)\, d\xi = 0 \quad \text{and}$$
(4.36)

$$\int_{-\infty}^{\infty} \{[A^{(0)}(\xi, \tau)]^2 - B^{(0)}(\xi, \tau)\}_\tau + \nu\{2[A^{(0)}(\xi, \tau)]^2 - B^{(0)}$$

$$(\xi, \tau)\}\, d\xi = 0,$$
(4.37)

obtained from integrating (4.13) and (4.14) with respect to ξ, respectively. It is straightforward to verify that (4.36) is satisfied for all $\tau \geq 0$. As is known rigorously, based on inverse scattering theory (Kaup and Newell 1978), slowly varying solitons evolve according to the energy balance relation. Consequently, in the present context, the slowly varying soliton *cannot* satisfy (4.37). Indeed, in comparing with (4.34), the left-hand side of (4.37) is exactly $-\mathrm{sgn}(c) \times$ (4.35). The shelf region develops behind the propagating soliton to compensate for the "mass" slowly lost or gained by the adiabatically modulating solitary wave.

The introduction of a slowly varying phase shift parameter into (4.16) will not eliminate the creation of the shelf region [or, equivalently, be chosen so that (4.37) will be satisfied for all $\tau \geq 0$]. Nor can the emergence of the shelf region be eliminated by the addition of a slowly varying contribution to $B^{(0)}$ [corresponding to a homogeneous solution to (4.14) at leading order]. The dynamics of the shelf region is determined by a more specialized analysis that is beyond the scope of this paper [Lamb (1971); see, also, Timko and Swaters (1997) for an analysis of dissipating internal solitons, or, more generally, Knickerbocker and Newell (1980) and Kodama and Ablowitz (1980)]. Physically, in the present context, the development of the shelf region in $B^{(1)}(\xi, \tau)$ corresponds to the emergence of an $O(\delta)$ slowly varying mean flow behind the solitary wave packet.

Last, the solution for $A^{(1)}(\xi, \tau)$ is given. Introducing the change of a dependent variable,

$$A^{(1)}(\xi, \tau) = A_\xi^{(0)}(\xi, \tau)\Phi(\xi, \tau),$$
(4.38)

into (4.32) leads to

$$(1 - c)(1 - c - 2k^2)\{[A_\xi^{(0)}]^2\Phi_\xi\}_\xi = -A_\xi^{(0)}[(1 - c)A_\xi^{(0)}]_\tau$$

$$-\frac{k^2 l^2 \Psi A^{(0)} A_\xi^{(0)}}{(1 - c)} - (1 - c - 2k^2)A_\xi^{(0)}A_{\xi\tau}^{(0)}$$

$$- 2\nu(1 - c - k^2)[A_\xi^{(0)}]^2,$$
(4.39)

which can be integrated twice to yield

$$\Phi = -\frac{(1 - c - 3k^2)\kappa^3\kappa_\tau\xi^2}{2\gamma}$$

$$+ \left[\frac{\mathrm{sgn}(c)(\gamma_\tau + \nu\gamma) - 3k^2\kappa^2\kappa_\tau(1 - c)\xi}{\gamma(1 - c)}\right]$$

$$\times [\kappa\xi - \xi\coth(\kappa\xi)] + \left[\frac{(1 - c - 2k^2)\gamma_\tau}{2k^2(1 - c)} - 3k^2\kappa\kappa_\tau\right.$$

$$\left. - 2\nu\kappa^2(1 - c - k^2)\right]\frac{\ln|\sinh(\kappa\xi)|}{\gamma},$$
(4.40)

which completes the solution. Observe that $A^{(1)} \to 0$ as $|\xi| \to \infty$ exponentially rapidly. There is no shelf region in $A^{(1)}$.

5. Summary and conclusions

The weakly nonlinear baroclinic instability characteristics of dissipative time-varying grounded abyssal flow have been described. Two situations are examined. The first corresponds to the weakly nonlinear *temporal* evolution of marginally stable or unstable abyssal flow that *does not* correspond to the point of marginal stability (i.e., the minimum baroclinic shear required for instability). It was shown that the resulting amplitude equations are identical (modulo a trivial rescaling) to the model derived by PT for the analogous problem for zonal baroclinic flow on a midlatitude β plane as described by the Phillips model. A very brief synopsis of the results was given.

The second situation examined corresponds to the weakly nonlinear and dissipative *spatial and temporal* (i.e., wave packet) evolution of marginally stable or unstable abyssal flow that *does* correspond to the point of marginal stability. It is known that, unlike the situation for modes associated with flows not located at the point of marginal stability, the dynamics in this case is fully nonlinear to leading order. Wave packet amplitude equations for dissipative time-varying flow are derived that are the analog of the time-only model derived by Warn and Gauthier (1989) for the Phillips model of the baroclinic instability of zonal flow on a β plane. These equations appear to be intractable.

A spectral decomposition is introduced that leads to an infinite set of coupled equations for the Fourier coefficients. If these equations are truncated, on an ad hoc basis, to include only the fundamental mode and its accompanying mean flow, the resulting model equations can be reduced to a perturbed sine–Gordon equation with time-varying coefficients. The SG equation has a soliton solution. A nonlinear WKB technique is introduced to determine the modulation of the soliton amplitude, wavenumber, and translation velocity when the time scale of the underlying flow variability is long relative to the advective time scale of the soliton with weak dissipation.

The evolution of the soliton parameters, when time variability and dissipation are present, satisfies an averaged energy balance equation. Detailed descriptions of the evolution of the soliton parameters were given when dissipation is present (but without time variability in the background marginal flow) and when time variability is present (but without dissipation).

The finite interval of allowed soliton translation velocities associated with a marginally *stable* background flow separates the allowed set of soliton translation velocities associated with a marginally *unstable* background flow into two disconnected semi-infinite intervals. The two values for the soliton velocity that separate the unstable and stable regions correspond to, respectively, the zero solution and a singular limit that can only exist if the background flow has no baroclinic shear.

It was shown that the dissipating soliton always evolves toward the zero solution. However, because of the structure of the solutions associated with the boundaries, with respect to the soliton translation velocity, between the unstable and stable regions, complex behavior can develop as the soliton is modulated as a result of dissipation or time variability. A solution was identified that can connect the marginally stable and unstable solitons if time variability in the back-

ground shear goes from being sub- to supercritical (and vice versa). These were illustrated with examples.

It was also shown that in the course of satisfying the averaged energy balance relation, the slowly deforming soliton is unable to satisfy the averaged mass balance relations. This results in the emergence of a small-amplitude "shelf region" in the nonlinearly induced mean flow that arises behind the propagating soliton. Last, the leading-order structure of the perturbation field associated with the modulated soliton, when dissipation and time variability are present, was determined.

There are, of course, several issues that remain outstanding. Perhaps most important, it would be interesting to solve the Warn and Gauthier (1989) formulation of the governing equations for the evolution of marginally stable or unstable flows at the point of marginal stability with time variability and/or spatial dependence retained (or both). In particular, it would be interesting to modify the solution procedure Warn and Gauthier (1989) developed for time-dependent-only solutions, to find, if they exist, solitary wave solutions. If this could be done, such a solution would correspond to the first genuinely coherent solitary wave solution, without artificial truncation, to a baroclinic instability problem for flows at minimum baroclinic shear, which is, after all, the most physically relevant transition problem.

REFERENCES

Ablowitz, M. J., and H. Segur, 1981: *Solitons and the Inverse Scattering Transform.* SIAM Press, 425 pp.

Benney, D. J., and R. F. Bergeron, 1969: A new class of nonlinear waves in parallel flows. *Stud. Appl. Math.,* **48,** 181–204.

Boville, B., 1981: Amplitude vacillation on a beta-plane. *J. Atmos. Sci.,* **38,** 609–618.

Bruce, J. G., 1995: Eddies southwest of Denmark Strait. *Deep-Sea Res.,* **42,** 13–29.

Charney, J. G., and G. R. Flierl, 1981: Oceanic analogues of large scale atmospheric motions. *Evolution of Physical Oceanography—Scientific Surveys in Honor of Henry Stommel,* B. A. Warren and C. Wunsch, Eds., MIT Press, 504–548.

de Verdiere, A. C., 1986: On mean flow instabilities within the planetary geostrophic equations. *J. Phys. Oceanogr.,* **16,** 1981–1984.

Drazin, P. G., and W. H. Reid, 1981: *Hydrodynamic Stability.* Cambridge University Press, 527 pp.

Etling, D., F. Gelhardt, U. Schrader, F. Brennecke, G. Kühn, G. Chabert d'Hieres, and H. Didelle, 2000: Experiments with density currents on a sloping bottom in a rotating fluid. *Dyn. Atmos. Oceans,* **31,** 139–164.

Flierl, G. R., 1984: Rossby wave radiation from a strongly nonlinear warm eddy. *J. Phys. Oceanogr.,* **14,** 47–58.

Gibbon, J. D., I. N. James, and I. M. Moroz, 1979: An example of soliton behavior in a rotating baroclinic fluid. *Proc. Roy. Soc. London,* **A367,** 219–237.

Grimshaw, R. H. J., 1979a: Slowly varying solitary waves, I. Korteweg–de Vries equation. *Proc. Roy. Soc. London,* **A368,** 359–375.

——, 1979b: Slowly varying solitary waves, II. Nonlinear Schroedinger equation. *Proc. Roy. Soc. London,* **A368,** 377–388.

Ha, S.-J., 2004: Finite-amplitude development of time dependent abyssal currents. M.S. thesis. Dept. of Mathematical and Statistical Sciences, University of Alberta, 101 pp.

Jiang, L., and R. W. Garwood Jr., 1996: Three-dimensional simulations of overflows on continental slopes. *J. Phys. Oceanogr.,* **26,** 1214–1233.

Jungclaus, J. H., J. Hauser, and R. H. Käse, 2001: Cyclogenesis in the Denmark Strait overflow plume. *J. Phys. Oceanogr.,* **31,** 3214–3229.

Kaup, D. J., and A. C. Newell, 1978: Solitons as particles and oscillators and in slowly-varying media: A singular perturbation theory. *Proc. Roy. Soc. London,* **A361,** 413–446.

Klein, P., and J. Pedlosky, 1992: The role of dissipation mechanisms in the nonlinear dynamics of unstable baroclinic waves. *J. Atmos. Sci.,* **49,** 29–48.

Knickerbocker, C. J., and A. C. Newell, 1980: Shelves and the Korteweg–de Vries equation. *J. Fluid Mech.,* **98,** 803–818.

Kodama, Y., and M. J. Ablowitz, 1980: Perturbations of solitons and solitary waves. *Stud. Appl. Math.,* **64,** 225–245.

Lamb, G. L., 1971: Analytical descriptions of ultrashort optical pulse propagation in a resonant medium. *Rev. Model. Phys.,* **43,** 99–124.

Mooney, C. J., and G. E. Swaters, 1996: Finite amplitude baroclinic instability of a mesoscale gravity current in a channel. *Geophys. Astrophys. Fluid Dyn.,* **82,** 173–205.

Nof, D., 1983: The translation of isolated cold eddies on a sloping bottom. *Deep-Sea Res.,* **30,** 171–182.

Pavec, M., X. Carton, and G. Swaters, 2005: Baroclinic instability of frontal geostrophic currents over a slope. *J. Phys. Oceanogr.,* **35,** 911–918.

Pedlosky, J., 1970: Finite amplitude baroclinic waves. *J. Atmos. Sci.,* **27,** 15–30.

——, 1972: Finite amplitude baroclinic wave packets. *J. Atmos. Sci.,* **29,** 680–686.

——, 1982a: Finite amplitude baroclinic waves at minimum shear. *J. Atmos. Sci.,* **39,** 555–562.

——, 1982b: A simple model for nonlinear critical layers in an unstable baroclinic wave. *J. Atmos. Sci.,* **39,** 2119–2127.

——, 1984: The equations for geostrophic flow in the ocean. *J. Phys. Oceanogr.,* **14,** 448–455.

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

——, and J. Thomson, 2003: Baroclinic instability of time dependent currents. *J. Fluid Mech.,* **490,** 189–215.

Poulin, F. J., and G. E. Swaters, 1999a: Sub-inertial dynamics of density-driven flows in a continuously stratified fluid on a sloping bottom. I. Model derivation and stability conditions. *Proc. Roy. Soc. London,* **A455,** 2281–2304.

——, and ——, 1999b: Sub-inertial dynamics of density-driven flows in a continuously stratified fluid on a sloping bottom. II. Isolated eddies and radiating cold domes. *Proc. Roy. Soc. London,* **A455,** 2305–2329.

——, and ——, 1999c: Sub-inertial dynamics of density-driven flows in a continuously stratified fluid on a sloping bottom. Part 3. Nonlinear stability theory. *Can. Appl. Math. Quart.,* **7,** 49–68.

Reszka, M. K., G. E. Swaters, and B. R. Sutherland, 2002: Instability of abyssal currents in a continuously stratified ocean with bottom topography. *J. Phys. Oceanogr.,* **32,** 3528–3550.

Samelson, R. M., and G. K. Vallis, 1997: A simple frictional and diffusion scheme for planetary geostrophic basin models. *J. Phys. Oceanogr.,* **27,** 186–194.

Swaters, G. E., 1991: On the baroclinic instability of cold-core coupled density fronts on sloping continental shelf. *J. Fluid Mech.,* **224,** 361–382.

——, 1993: Nonlinear stability of intermediate baroclinic flow on a sloping bottom. *Proc. Roy. Soc. London,* **A442,** 249–272.

——, 1998a: Numerical simulations of the baroclinic dynamics of density-driven coupled fronts and eddies on a sloping bottom. *J. Geophys. Res.,* **103,** 2945–2961.

——, 1998b: Dynamics of radiating cold domes on a sloping bottom. *J. Fluid Mech.,* **364,** 221–251.

——, and G. R. Flierl, 1991: Dynamics of ventilated coherent cold eddies on a sloping bottom. *J. Fluid Mech.,* **223,** 565–587.

Tan, B., and J. P. Boyd, 2002: Envelope solitary waves and periodic waves in the AB equations. *Stud. Appl. Math.,* **109,** 67–87.

Timko, P. G., and G. E. Swaters, 1997: On the dissipation of internal solitons in coastal seas. *Geophys. Astrophys. Fluid Dyn.,* **85,** 163–194.

Warn, T., and P. Gauthier, 1989: Potential vorticity mixing by marginally unstable waves at minimum shear. *Tellus,* **41A,** 115–131.