PROCEEDINGS OF THE ROYAL SOCIETY A

Proc. R. Soc. A (2007) **463**, 1749–1769
doi:10.1098/rspa.2007.1850
Published online 24 April 2007

Evolution of solitary marginal disturbances in baroclinic frontal geostrophic dynamics with dissipation and time-varying background flow

By Mattea R. Turnbull and Gordon E. Swaters*

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G2G1

A two-layer frontal geostrophic flow corresponds to a dynamical regime that describes the low-frequency evolution of baroclinic ocean currents with large amplitude deflections of the interface between the layers on length-scales longer than the internal deformation radius within the context of a thin upper layer overlying a dynamically active lower layer. The finite-amplitude evolution of solitary disturbances in baroclinic frontal geostrophic dynamics in the presence of time-varying background flow and dissipation is shown to be governed by a two-equation extension of the *unstable* nonlinear Schrödinger (UNS) equation with variable coefficients and forcing. The soliton solution of the unperturbed UNS equation corresponds to a saturated isolated coherent anomaly in the baroclinic instability of surface-intensified oceanographic fronts and currents. The adiabatic evolution of the propagating soliton and the uniformly valid first-order perturbation fields are determined using a direct perturbation approach together with phase-averaged conservation relations when both dissipation and time variability are present. It is shown that the soliton amplitude parameter decays exponentially due to the presence of the dissipation but is unaffected by the time variability in the background flow. On the other hand, the soliton translation velocity is unaffected by the dissipation and evolves only in response to the time variability in the background flow. The adiabatic solution for the induced mean flow exhibits a dissipation-generated 'shelf region' in the far field behind the soliton, which is removed by solving the initial-value problem.

Keywords: frontal geostrophic dynamics; ocean eddies and fronts; baroclinic instability; unstable nonlinear Schrödinger equation; perturbations of solitons

1. Introduction

An important kinematic feature of surface-intensified ocean currents is the fact that the dynamic deflections of the surfaces of constant density, called isopycnals or the geopotentials, have the same order of magnitude as the depth of the fluid itself. This property implies that these flows cannot be modelled with classical quasigeostrophic (QG) theory (Pedlosky 1987), which requires that the amplitude of the dynamic deflections of the geopotentials is small when compared with the mean depth of the fluid.

* Author for correspondence (gordon.swaters@ualberta.ca).

Received 3 February 2007
Accepted 30 March 2007

Frontal geostrophic (FG) models (Cushman-Roisin 1986; Cushman-Roisin *et al.* 1992; Swaters 1993; Karsten & Swaters 2000) correspond to a dynamical regime that describes stratified sub-inertial ocean currents with large amplitude isopycnal deflections on length-scales longer than the internal deformation radius (approx. 100 km, this is the length-scale at which Earth's rotation can no longer be neglected in the dynamics). In particular, FG theory can describe surface ocean currents in which the isopycnals intersect the surface (i.e. outcroppings), as they do in frontal regions and warm-core rings and eddies.

Frontal geostrophic dynamics allows for the 'vigorous' baroclinic instability (i.e. the transfer of mean flow potential energy to perturbation kinetic energy) in frontal regions and the subsequent formation of isolated coherent eddies (e.g. Tang & Cushman-Roisin 1992; Reszka & Swaters 1999*a*,*b*, 2004). Thus, FG dynamics provides a model for the formation, transport and mixing over large distances of isolated mesoscale anomalies, which are ubiquitous features of turbulent geostrophic flow (McWilliams 1984).

Swaters (1993) showed that the two-layer FG equations can be written as an infinite-dimensional, non-canonical Hamiltonian dynamical system (Swaters 2000) and used this formalism to establish sufficient linear and nonlinear stability and necessary instability conditions for the steady flow solutions of the model. Reszka & Swaters (1999*a*) described the linear and nonlinear baroclinic instability characteristics and presented fully nonlinear numerical simulations for various baroclinic flows described by this model. When there are no non-conservative processes present and the background flow is steady, in the weakly nonlinear and marginally stable or unstable regime, the leading-order disturbance amplitude satisfies the *unstable* nonlinear Schrödinger (UNS) equation. (For the derivation of the UNS equation in other contexts, see Gibbon & McGuinness 1981; Yajima & Tanaka 1988; Iizuka & Wadati 1990; Yajima & Wadati 1990*b*; Tan & Liu 1995. More will be said about the UNS equation later.) The numerical simulations described by Reszka & Swaters (1999*a*,*b*, 2004) show the formation of isolated surface-intensified 'warm-core', or anticyclonic, rings and eddies, whose boundaries correspond to an outcropping (i.e. the eddies were compactly supported in the mathematical sense of the term) as a consequence of baroclinic destabilization. The numerical simulations were also able to show that the FG baroclinic model can describe eddy–eddy and eddy–mean flow interactions for vortices with outcroppings. These are oceanographic processes beyond the ability of QG theory to describe.

Recent work has shown that time variability in the background flow and dissipative processes within the fluid can have a profound effect on the *linear and nonlinear* stability characteristics of atmospheric and ocean currents (Klein & Pedlosky 1992; Pedlosky & Thomson 2003; Poulin *et al.* 2003; Ha & Swaters 2006). Indeed, even if the time average of the background flow is itself stable, small amplitude oscillations can lead to linear destabilization or vice versa (even if the oscillatory flow is, at each moment in time, linearly stable or unstable, respectively).

This is a very important observation because it means that there are entirely new transitions to (or the suppression of) instability sequences that may be occurring in the ocean or atmosphere. For example, the periodic perturbing of a coastal current by the tides may very well lead to destabilization even if the coastal current is instantaneously stable at each moment in time. On the other hand, and in the other direction, it is possible that long time-scale periodic, e.g. seasonal, forcing may act to stabilize an otherwise unstable large-scale current. The implications of these types of processes on the planetary-scale mixing of the oceans may make an important contribution to climate variability.

The principal purpose of this paper has been to develop a theory for the weakly nonlinear evolution of marginally stable or unstable baroclinic FG flow when dissipative processes and a time-varying background current are present. It is shown that the governing equations describing the leading-order wave-packet dynamics correspond to an extended two-equation form of the UNS model with variable coefficients and forcing.

Two important sub-limits of the model equations have already been examined, albeit obtained in other contexts. In the case when spatial variations are neglected and there is no background time variability (but dissipation is present), the system reduces to the Lorenz equations, in which the Prandtl number is exactly 1. When background time variability is present (but dissipation and spatial variations are neglected), the model is equivalent to the 'AB equations' (without spatial derivatives) that govern some aspects of finite-amplitude baroclinic instability in the QG and other approximations (e.g. Pedlosky 1972; Gibbon *et al.* 1979; Gibbon & McGuinness 1981; Mooney & Swaters 1996). These sub-limits have previously been examined by Klein & Pedlosky (1992), Pedlosky & Thomson (2003) and Ha & Swaters (2006).

However, the goal here will be to focus on the situation where spatial dependence is essential. In particular, the dynamical characteristics of the soliton solutions to the extended UNS system derived here are determined when dissipation and a time-varying background flow are present, in an effort to understand the modulational properties of the coherent solitary structures that can emerge in the transition to instability in baroclinic FG dynamics.

2. Derivation of the wave-packet model

(*a*) *The frontal geostrophic model*

A detailed description of the mathematical derivation of FG models can be found in Cushman-Roisin (1986), Cushman-Roisin *et al.* (1992), Swaters (1993) and Karsten & Swaters (2000). Briefly, baroclinic FG models correspond to a sub-inertial asymptotic limit of the two-layer shallow water equations (the theory can be extended to multi-layer, continuously stratified fluids), in which the flow fields are geostrophic to leading order but for which ageostrophic effects must be included to determine the evolution. FG models allow for large amplitude variations of the upper layer thickness so that genuine outcroppings can be described. Unlike the classical QG limit, the upper layer mass equation is *fully nonlinear* and does not reduce to the statement that the velocity is solenoidal to leading order. The dynamics in the lower layer is 'driven' by the vortex stretching associated with the deforming interface and a background vorticity gradient, i.e. the β-effect or variable bottom topography.

The two-layer FG equations with variable bottom topography, Rayleigh dissipation and a total volume-conserving, interlayer mass exchange term (that is needed to drive forced time-dependent upper layer flow) can be obtained from

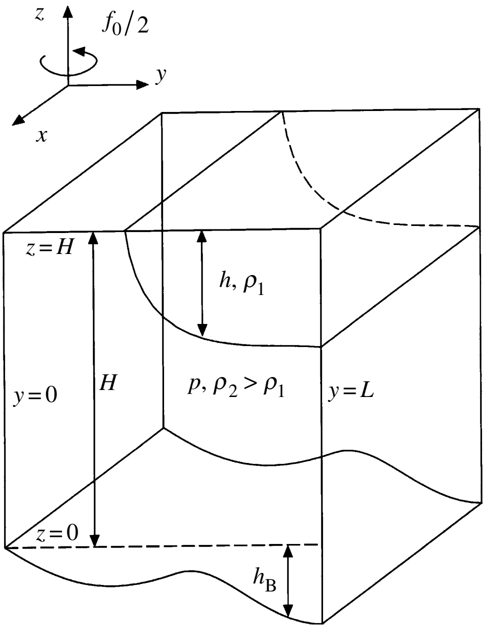

Figure 1. Physical configuration of the two-layer frontal geostrophic model. The fluid is stably stratified and rotating about the z-axis with constant angular frequency $f_0/2$. The domain corresponds to a channel, with variable bottom topography, oriented parallel to the x-axis with fixed side walls located at $y=0$ and L, respectively. The thickness of the upper layer can have large amplitude deflections with the possibility that the layer thickness can dynamically vanish along some curve, say $y=\hat{y}(x,t)$, (i.e. an outcropping can form).

the non-dimensional, two-layer, rigid-lid, shallow water equations written in the form

$$\delta(\delta\partial_t + \boldsymbol{u}_1\cdot\nabla)\boldsymbol{u}_1 + \boldsymbol{e}_3 \times \boldsymbol{u}_1 = -\nabla p_1 - \delta^2\nu\boldsymbol{u}_1, \tag{2.1}$$

$$\delta h_t + \nabla\cdot(\boldsymbol{u}_1 h) = \delta(F - \nu h), \tag{2.2}$$

$$\delta^2(\partial_t + \boldsymbol{u}_2\cdot\nabla)\boldsymbol{u}_2 + \boldsymbol{e}_3 \times \boldsymbol{u}_2 = -\nabla p_2 - \delta^2\nu\boldsymbol{u}_2, \tag{2.3}$$

$$\nabla\cdot\boldsymbol{u}_2 = \delta^2\{h_t + \nabla\cdot[\boldsymbol{u}_2(h+h_B)] - F + \nu h\}, \tag{2.4}$$

$$p_1 = h + \delta p_2, \tag{2.5}$$

where the horizontal coordinates are (x, y); t is the time; $\nabla=(\partial_x, \partial_y)$, where the alphabetical subscripts indicate partial differentiation (unless otherwise indicated) and the subscripts 1 and 2 denote, respectively, the upper and lower layer quantities, with $\boldsymbol{u}_{1,2}$, $p_{1,2}$, h and h_B the velocities, reduced pressures, upper layer thickness and the height of the bottom topography, respectively; and $\delta^2 \equiv h_*/H$ is the ratio of the upper layer scale thickness h_* to the lower layer scale thickness H (figure 1). Written in this form, the key asymptotic parameter is δ for which it is assumed that $0<\delta\ll 1$ (note that necessarily $\delta<1$, figure 1).

The Rayleigh dissipation coefficient is ν (assumed, for convenience, to be the same in each layer) and $F(x, y, t)$ is the interlayer mass exchange term needed to drive forced time-dependent upper layer flow. The physical motivation for the above modelling of the dissipation and interlayer mass exchange is that this formulation ensures that the dissipation is proportional to the leading-order dynamical potential vorticity (PV, see the discussion following equations (2.6) and (2.7) below) and that the mass exchange does not lead to a non-zero divergence in the barotropic mass flux (Klein & Pedlosky 1992; Swaters 2006), i.e. it follows from (2.2) and (2.4) that

$$\nabla\cdot\{\delta \boldsymbol{u}_1 h + \boldsymbol{u}_2[1+\delta^2(h+h_{\mathrm{B}})]\} = 0.$$

The time and length scales used to derive (2.1)–(2.5) are given by

$$T = \frac{1}{f_0\delta^2} \quad \text{and} \quad L = \frac{\sqrt{g' h_*/\delta}}{f_0},$$

respectively, where f_0 and $g' = (\rho_2 - \rho_1)g/\rho_2 > 0$ are the Coriolis parameter and reduced gravity, respectively. The scalings for the upper and lower layer velocities are given by

$$U_1 = \delta f_0 L \quad \text{and} \quad U_2 = \delta^2 f_0 L,$$

respectively, and the reduced pressures $p_{1,2}$ are scaled geostrophically.

The baroclinic FG model is obtained by introducing a straightforward asymptotic expansion in the powers of δ into the above shallow water equations (the details of which are not given here, see Swaters 1993). If the leading-order upper layer frontal thickness is denoted by h, which is also the leading-order geostrophic pressure in the upper layer (i.e. $p_1 = h$), and the leading-order lower layer geostrophic pressure is denoted by $p_2 \equiv p$, the FG model can be written in the form

$$h_t + J\left(p + h\Delta h + \frac{\nabla h\cdot\nabla h}{2}, h\right) = -\nu h + F, \tag{2.6}$$

$$(\Delta p + h)_t + J(p, \Delta p + h + h_{\mathrm{B}}) = -\nu(\Delta p + h) + F, \tag{2.7}$$

where $J(A, B) = A_x B_y - A_y B_x$. The leading-order Eulerian velocity fields will be given by

$$\boldsymbol{u}_1 = \boldsymbol{e}_3 \times \nabla h \quad \text{and} \quad \boldsymbol{u}_2 = \boldsymbol{e}_3 \times \nabla p.$$

The leading-order dynamic PV in the upper and lower layers is given by $1/h$ and $\Delta p + h$, respectively. Equations (2.6) and (2.7) are the PV equations for the two layers. The dissipation terms are proportional to the individual layer dynamical PVs (Klein & Pedlosky 1992).

Equation (2.6) permits solutions which possess the property that h can intersect the surface (i.e. outcrop) along dynamically evolving curve(s), generically denoted by, say, $y = \hat{y}(x, t)$. Swaters (1993) has shown that equation (2.6) *is itself* the appropriate kinematic and dynamic boundary conditions for h on the outcropping. This means that when an outcropping is present, the solution of equation (2.6) automatically determines the correct evolution and placement of the outcropping (note that $h \equiv 0$ solves equation (2.6)). No additional matching, continuity or boundary conditions for h are required at or across the outcropping or are needed to determine, formally, $y = \hat{y}(x, t)$. This is a noteworthy feature of the two-layer FG model (2.6) and (2.7).

The underlying non-canonical Hamiltonian structure and the linear and nonlinear stability theory associated with steady solutions to equations (2.6) and (2.7) are described by Swaters (1993), Karsten & Swaters (1996, 2000) and Reszka & Swaters (1999*a*). Of particular importance is the fact that all steady solutions to the inviscid unforced barotropic problem (i.e. $\nu = F = p = 0$, with equation (2.7) ignored) are unconditionally linearly stable (Swaters 1993). Thus, baroclinic coupling is necessary for instability. Phenomenologically, this is important because it means that the unstable modes are not simply baroclinically modified barotropic instabilities, but they represent genuine baroclinic FG instabilities.

(*b*) *The amplitude equations*

The amplitude equations are obtained by considering the weakly nonlinear slow evolution of a marginally stable or unstable surface-intensified flow over gently linearly sloping topography (Reszka & Swaters 1999*a*; Turnbull 2006). To that end, $h_B = -s(y - L/2)$, where $0 < s \ll 1$ is the topographic slope and $y \in (0, L)$. The flow occurs in a channel of width L, with x and y the along- and cross-channel coordinates, respectively. The sloping topography gives rise to a background vorticity gradient that provides a waveguide for the normal modes, which are topographic Rossby waves. The linear destabilization corresponds to the coalescence of the barotropic and baroclinic topographic Rossby wave modes.

This can be seen by determining the normal mode instabilities associated with the inviscid unforced 'wedge front' solution of equations (2.6) and (2.7) given by

$$\nu = F = p = 0 \quad \text{and} \quad h = h_0(y) = 1 + \alpha s(y - L/2), \tag{2.8}$$

where αs is (for now) the constant slope of the interface between the upper and lower layers. The leading order (with respect to s) perturbed normal mode solution to equations (2.6) and (2.7), satisfying $v_{1,2} = 0$ on the channel walls, can be written in the form (Reszka & Swaters 1999*a*; Turnbull 2006)

$$(h, p) = (h_0, 0) + \{(A, B)\sin(ly)\exp[\mathrm{i}k(x - c\tilde{t})] + \text{c.c.}\}, \tag{2.9}$$

where $l \equiv n\pi/L$ ($n \in \mathbb{Z}^+$); $\tilde{t} = st$; k is the real-valued along channel wavenumber; $c = c_R + ic_I$ is the complex-valued phase velocity, $\mathrm{i}^2 = -1$; c.c. is the complex conjugate of the previous term and where

$$B = [K^2 + (\alpha - 1)/c]A \quad \text{where } K^2 \equiv k^2 + l^2, \tag{2.10}$$

with the dispersion relation

$$c = \frac{1 - \alpha K^4 \pm [(1 - \alpha K^4)^2 - 4\alpha(\alpha - 1)K^4]^{1/2}}{2K^2}. \tag{2.11}$$

The introduction of the time variable $\tilde{t}$ and the assumption that the mean interface slope is $O(s)$ have been made since the phase or group velocity of a Rossby wave is proportional to the background vorticity gradient (planetary or topographic) and to ensure that the baroclinic vortex stretching associated with the deformed mean interface is the same order of magnitude as the stretching associated with the topographic slope.

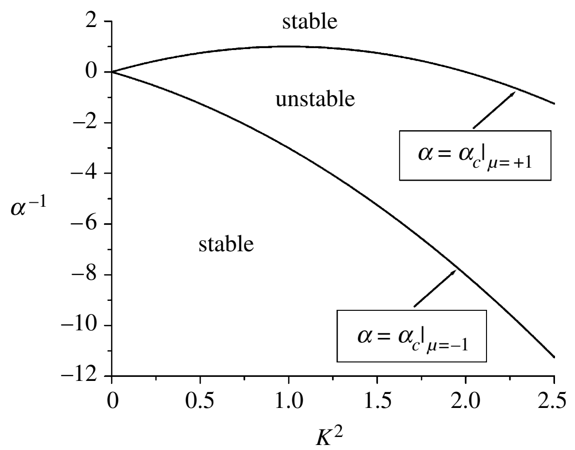

Figure 2. Stability diagram in the (K^2, α^{-1})-plane for the dispersion relation (2.11). The two stability boundaries are given by $\alpha = \alpha_c \equiv \mu/[K^2(2-\mu K^2)]$, with $\mu = +1$ and -1 corresponding to the 'upper' and 'lower' curves, respectively. The region bounded by the boundary curves is unstable and the remaining two 'outer' regions are stable. The wave-packet model derived here corresponds to examining the finite-amplitude dissipative evolution of a normal mode located 'near' the stability boundaries with time variability included in the background flow.

Instability develops when the discriminate in equation (2.11) is negative, which in terms of α occurs when

$$-K^2(2+K^2) < \alpha^{-1} < K^2(2-K^2). \tag{2.12}$$

In terms of α, there are two marginal stability boundaries, which can be expressed in the form

$$\alpha = \alpha_c \equiv \frac{\mu}{K^2(2-\mu K^2)}, \tag{2.13}$$

where $\mu = +1$ and -1 corresponds to the upper and lower boundaries in equation (2.12), respectively. It follows from equations (2.10) and (2.11) that

$$B = \mu A \quad \text{and} \quad c = \frac{(1-\mu K^2)}{K^2(2-\mu K^2)},$$

when $\alpha = \alpha_c$. Thus, the $\mu = +1$ and -1 marginal stability boundaries correspond to the barotropic and baroclinic modes, respectively. The region of instability and the stability boundaries in the (K^2, α^{-1})-plane are shown in figure 2.

It follows from equation (2.12) that a mode will be located near the marginal stability boundaries when $\alpha = \alpha_c \pm \mu\Delta^2$, where $0 < \Delta^2 \ll 1$ and the '+' and '−' signs denote a marginally unstable or stable mode, respectively. Additionally, it follows from equation (2.11) that the growth rate $\sigma \equiv kc_{\mathrm{I}} \simeq O(|\Delta|) > 0$ in the marginally unstable case (i.e. associated with the '+' sign; $c_{\mathrm{I}} = 0$ in the marginally stable situation).

It is now possible to briefly but precisely formulate the asymptotic regime for the finite-amplitude wave-packet model. The marginal time-dependent background flow is given by

$$p = 0 \quad \text{and} \quad h = h_0 = 1 + \alpha s(y - L/2), \tag{2.14}$$

with

$$\alpha = \alpha_c \pm \mu s^2 \Phi(T) \quad \text{where } T = s^2 t, \tag{2.15}$$

where $\Phi(T) \simeq O(1)$ is a real-valued function describing the slow time evolution of the background flow. The flow equations (2.14) and (2.15) will be a solution of equations (2.6) and (2.7) if

$$F \equiv (\partial_t + \nu) h_0 = s^2 (\partial_T + \tilde{\nu}) h_0,$$

where $\nu = s^2 \tilde{\nu}$ has been introduced to ensure that the dissipation will contribute to the dynamics at the same order of magnitude as the nonlinear interactions. To account for slow space development, the long scale $X = s^2 x$ is introduced. It can be shown that there is no dynamically relevant 'shorter' long space scale in the weakly nonlinear marginal problem (Reszka 1997).

With these assumptions, if one constructs an asymptotic solution to equations (2.6) and (2.7) in the form

$$(h, p) = (h_0, 0) + s^2\{(1, \mu) A \sin(ly) \exp[\mathrm{i}k(x - c\tilde{t})] + \text{c.c.}\} + O(s^3), \tag{2.16}$$

it follows after substantial algebra (one must actually proceed to the $O(s^4)$ problem to complete the analysis; see Reszka 1997; Reszka & Swaters 1999*a*; Turnbull 2006) that

$$[(\partial_T + \nu)^2 - \sigma(T)] A - \mathrm{i} P A_X + NAB = 0, \tag{2.17}$$

$$(\partial_T + \nu) B = (\partial_T + 2\nu) |A|^2, \tag{2.18}$$

where $A(X, T)$ is the complex-valued modal amplitude in (2.16); $B(X, T)$ is the leading-order, real-valued mean flow that is created by the self-interaction of the fundamental with itself; and the tilde on ν has been deleted.

In equations (2.17) and (2.18), P and N are constants given by

$$P \equiv \frac{4\mu k^3 (K^2 - \mu)}{K^6 (K^2 - 2\mu)^2}, \quad N \equiv (lk)^2 \left[2 - \mu K^2 + \frac{4\mu l^2 (K^2 - \mu)}{K^2} \right], \tag{2.19}$$

where σ is given by

$$\sigma(T) \equiv \frac{k^2 \Phi(T)}{K^2} + \frac{2\alpha_c k^2 (K^2 - \mu)}{L K^6 (K^2 - 2\mu)^2} \int_0^L [\tilde{h}_y + (y - L/2)(\tilde{h}_{yy} - k^2 \tilde{h}) \sin(ly)] \, \mathrm{d}y, \tag{2.20}$$

where

$$\tilde{h} \equiv (\partial_{yy} + l^2 - \mu) \tilde{p}, \tag{2.21}$$

with $\tilde{p}$ the solution of

$$(\partial_{yy} + l^2 - 2\mu)(\partial_{yy} + l^2) \tilde{p} = \alpha_c \mu [l \cos(ly) - K^2 (y - L/2) \sin(ly)]. \tag{2.22}$$

The solution for $\tilde{p}$ and the subsequent evaluation of $\tilde{h}$ and the integral in equation (2.20) are straightforward but dependent on the sign of $l^2 - 2\mu$. The precise form is not material for the presentation here and so is not included in the interests of conserving space. Full details can be found in Reszka (1997), Reszka & Swaters (1999*a*) and Turnbull (2006).

The parameter N in equation (2.17) is the coefficient of the nonlinear term. If $N=0$, the mode evolves linearly to this order in the theory. In what follows, attention is restricted to the stable situation for which $N>0$. The case where $N<0$ is 'explosively unstable' (e.g. Craik 1985), in that the T-only solutions become unbounded in finite time for all initial conditions if the flow is linearly unstable or when the initial amplitude is sufficiently large if the flow is linearly stable (Turnbull 2006).

The expression for $\sigma(T)$ contains two terms. The first term on the r.h.s of equation (2.20) is the slow time variability of the background flow that arises due to the $O(s^2)$ sub- or super-criticality in equation (2.15). The second term on the r.h.s. of equation (2.20) (the integral term) arises from additional linear terms in the higher-order perturbation problems that are not present in the leading-order equation.

Finally, the wave-packet model in the form convenient for the present purpose is obtained by introducing the transformation

$$A=\sqrt{2/N}\tilde{A}(\tilde{X},T),\quad B=(2/N)[|A|^2(\tilde{X},T)+\tilde{B}(\tilde{X},T)],\quad \text{where}\quad \tilde{X}=X/P,$$

into equations (2.17) and (2.18), which yields (after dropping the tildes)

$$[(\partial_T+\nu)^2-\sigma]A-\mathrm{i}A_X+2(|A|^2+B)A=0, \tag{2.23}$$

$$(\partial_T+\nu)B=\nu|A|^2. \tag{2.24}$$

The model (2.23) and (2.24) is a two-equation extension of the UNS equation. The canonical form of the UNS equation corresponds to the inviscid $\nu=B=0$ (assuming σ, i.e. Φ, is constant) limit of equations (2.23) and (2.24), which can be written in the form

$$q_{TT}-\mathrm{i}q_X+2|q|^2q=0,$$

where $q=A\exp(-\mathrm{i}\sigma X)$. The UNS equation is a special case of the Ginzburg–Landau equation and is identical in form to the nonlinear Schrödinger (NLS) equation with the space and time variables interchanged. The UNS equation has been derived in a number of other contexts (Gibbon & McGuinness 1981; Yajima & Tanaka 1988; Iizuka & Wadati 1990; Yajima & Wadati 1990*b*; Tan & Liu 1995) and is a 'canonical' integrable model describing the weakly nonlinear evolution of marginal unstable disturbances in a dispersive medium. The inverse scattering transform for the UNS equation was described by Yajima & Wadati (1990*a*).

Before moving on to describe the evolution of the soliton solution to equations (2.23) and (2.24) when dissipation and time variability in the background flow are present, it is useful to very briefly describe the qualitative properties of this system when the spatial derivative is neglected. In this situation, the appropriate initial conditions for the marginally unstable problem are $A(0)=A_0$, $A_T(0)=k\sqrt{|\sigma(0)|}A_0/K$ and $B(0)=-|A_0|^2$ (i.e. initially, there is no mean flow and the modal amplitude increases at the rate given by the leading-order growth rate). In this case, the solution for $B(T)$ is simply

$$B(T)=\mathrm{e}^{-\nu T}\left[\nu\int_0^T \mathrm{e}^{\nu\xi}|A|^2(\xi)\mathrm{d}\xi-|A_0|^2\right],$$

and equation (2.23) can be written in the form

$$[(\partial_T+\nu)^2-\sigma]A+2\left(|A|^2+\mathrm{e}^{-\nu T}\left[\nu\int_0^T \mathrm{e}^{\nu\xi}|A|^2(\xi)\mathrm{d}\xi-|A_0|^2\right]\right)A=0. \tag{2.25}$$

If the background flow does not possess time variability (i.e. $\Phi(T)$ and hence σ is constant), then equation (2.25) is a Lorenz dynamical system in which the Prandtl number is 1 (Klein & Pedlosky 1992). With dissipation present, 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 is the motivation for modelling the dissipation in equations (2.6) and (2.7) as proportional to the dynamical PV (Klein & Pedlosky 1992).

Time variability in the background flow can have a profound influence on the stability properties (e.g. Pedlosky & Thomson 2003; Poulin *et al.* 2003) of atmospheric and oceanographic currents. For example, even if the flow is marginally *stable* ($\sigma<0$), an oscillatory component (e.g. $\Phi(T)=\Phi_0\cos(\omega T)$) does, generically in the inviscid limit, linearly destabilize the flow (Pedlosky & Thomson 2003). However, if the flow is marginally *unstable* ($\sigma>0$), an oscillatory component will not, generically in the inviscid limit, stabilize the flow. (Notwithstanding this generic behaviour, it is possible to have an oscillatory component that will stabilize, even in the linear inviscid problem, an otherwise unstable flow.) These important new results suggest that the transition to instability can occur for a larger, and more importantly, more realistic range of flow parameters. Regardless, the nonlinear terms in equation (2.25) act, ultimately, to bound the evolution of $A(T)$ (with or without dissipation being present).

Finally, it is important to be reminded that equations (2.23) and (2.24) are *envelope equations* that describe the slow temporal and long spatial evolution of the *amplitude* of the fast phase oscillations associated with the carrier wave in equation (2.9). Consequently, implicit in the derivation of equations (2.23) and (2.24) is the assumption that there is a length-scale separation between carrier wave and the solutions of equations (2.23) and (2.24). For this to be maintained, it is required that the coefficient P in equation (2.17) satisfies $P\sim O(1)$. Generically, it follows from equation (2.19) that this condition will only fail to hold, i.e. $|P|\gg 1$, for the barotropic mode (i.e. $\mu=+1$) when $|K^2-2\mu|\ll 1$, i.e. when $K\simeq\sqrt{2}$. For the baroclinic mode (i.e. $\mu=-1$), $|K^2-2\mu|>2$, so that $P\sim O(1)$ for all K since $K>\pi/L>0$.

3. Soliton evolution with dissipation and time variability

Huang *et al.* (2000) have presented a perturbation calculation for the UNS equation based on the inverse scattering theory of Yajima & Wadati (1990*a*). Unfortunately, the Huang *et al.* calculation appears to be in error (at least in as much as how it was applied for a dissipative perturbation). The correct theory for the adiabatic evolution of the perturbed soliton solution to the UNS equation, derived independently based on both direct perturbation and fast phase-averaged conservation relation approaches (for these techniques applied to other soliton problems, see Grimshaw 1979*a*,*b*; Kaup & Newell 1978; Kodama & Ablowitz 1980), has been described by Swaters (2007). Here, these methods are applied to determine the evolution of the soliton solution to equations (2.23) and (2.24) in the weakly dissipative limit, where the background time variability in the mean flow varies slowly.

When $\nu = 0$ and σ is a constant, the soliton solution to equations (2.23) and (2.24) can be written in the form

$$A_{\text{soliton}}(X,T) = \mu\,\text{sech}\left[\frac{\mu(X - X_0 - UT)}{U}\right] \times \exp\left[\text{i}\left\{\left(\sigma + \mu^2 - \frac{1}{4U^2}\right)(X - X_0 - UT) + U\left(\sigma + \mu^2 + \frac{1}{4U^2}\right)(T - T_0)\right\}\right], \tag{3.1}$$

$$B_{\text{soliton}}(X,T) \equiv 0, \tag{3.2}$$

where μ, U and (X_0, T_0) are the arbitrary real-valued amplitude, translation velocity and space–time phase-shift parameters, respectively. The goal here is to determine the adiabatic evolution of the soliton when there is weak dissipation and slowly varying time variability in the background flow, that is, to determine the leading-order solution to equations (2.23) and (2.24), assuming

$$A(X,0) = A_{\text{soliton}}(X,0), \quad A_T(X,0) = \partial_T A_{\text{soliton}}(X,0) \quad \text{and} \quad B(X,0) = 0, \tag{3.3}$$

to leading order.

To that end, it is assumed that

$$\nu = \varepsilon\hat{\nu} \quad \text{and} \quad \sigma = \sigma(\varepsilon t) \quad \text{where} \quad 0 < \varepsilon \ll 1, \tag{3.4}$$

which motivates the introduction of the fast phase and slow time variables

$$\theta = x - \frac{1}{\varepsilon}\int_0^{\varepsilon T} U(\xi)\,\text{d}\xi \quad \text{and} \quad \tau = \varepsilon T, \tag{3.5}$$

respectively, and

$$\Phi \equiv \frac{1}{\varepsilon}\int_0^{\varepsilon T} U(\xi)\left[\sigma(\xi) + \mu^2(\xi) + \frac{1}{4U^2(\xi)}\right]\text{d}\xi, \tag{3.6}$$

so that

$$\theta_T = -U(\tau), \quad \Phi_T = U(\tau)\left[\sigma(\tau) + \mu^2(\tau) + \frac{1}{4U^2(\tau)}\right], \tag{3.7}$$

$$\partial_X \to \partial_\theta \quad \text{and} \quad \partial_{TT} \to U^2\partial_{\theta\theta} - \varepsilon[U\partial_{\theta\tau} + (U\partial_\theta)_\tau] + \varepsilon^2\partial_{\tau\tau}. \tag{3.8}$$

In terms of θ and τ, the soliton solution (3.1) can be written in the form

$$A_{\text{soliton}} = \mu\,\text{sech}(\mu\theta/U)\exp\left[\text{i}\left\{\left(\sigma + \mu^2 - \frac{1}{4U^2}\right)\theta + \Phi\right\}\right]. \tag{3.9}$$

In equation (3.9) it has been assumed that $X_0 = T_0 = 0$. As in the perturbed NLS or KdV problems (Kaup & Newell 1978; Grimshaw 1979*a*,*b*; Kodama & Ablowitz 1980), it can be shown (Swaters 2007) that the leading-order solvability conditions, or equivalently, phase-averaged conservation relations do not determine the evolution of the phase-shift parameters and thus, without loss of generality for the analysis presented here, they may be set to zero. Their evolution is determined by higher-order solvability conditions or, equivalently, higher-order fast phase-averaged conservation relations, which are not examined here.

The solution for the adiabatically evolving soliton is most conveniently obtained in the form

$$A(\theta,\tau) = [\eta(\theta,\tau) + \mathrm{i}\varepsilon\phi(\theta,\tau)]\exp\left[\mathrm{i}\left\{\left(\sigma+\mu^2-\frac{1}{4U^2}\right)\theta+\Phi\right\}\right], \tag{3.10}$$

$$B(\theta,\tau) = \varepsilon\psi(\theta,\tau), \tag{3.11}$$

where $\eta(\theta,\tau)$, $\phi(\theta,\tau)$ and $\psi(\theta,\tau)$ are real-valued. Substitution of equations (3.4)–(3.6), (3.10) and (3.11) into equations (2.23) and (2.24) leads to, after some algebra (including dropping the caret on $\hat{\nu}$ and separating out the real and imaginary parts), respectively,

$$(U^2\partial_{\theta\theta}-\mu^2)\eta + 2\eta^3 = \varepsilon F_{\mathrm{R}}(\eta,\psi) + O(\varepsilon^2), \tag{3.12}$$

$$(U^2\partial_{\theta\theta}-\mu^2+2\eta^2)\phi = F_{\mathrm{I}}(\eta) + O(\varepsilon), \tag{3.13}$$

$$U\psi_\theta + \nu\eta^2 = \varepsilon(\partial_\tau+\nu)\psi + O(\varepsilon^2), \tag{3.14}$$

where

$$F_{\mathrm{R}}(\eta,\psi) \equiv 2\nu U\eta_\theta - 2\psi\eta + (U\eta_\theta)_\tau + U\eta_{\theta\tau} + (\theta\eta/U)\left(\sigma+\mu^2-\frac{1}{4U^2}\right)_\tau, \tag{3.15}$$

$$F_{\mathrm{I}}(\eta) \equiv -\nu\eta/U - \eta_\tau/U + 2\theta\eta_\theta U\left(\sigma+\mu^2-\frac{1}{4U^2}\right)_\tau + \eta\left[U\left(\sigma+\mu^2-\frac{1}{4U^2}\right)_\tau + \frac{U_\tau}{2U^2}\right]. \tag{3.16}$$

Substitution of the straightforward expansion

$$(\eta,\phi,\psi) \simeq (\eta,\phi,\psi)^{(0)} + \varepsilon(\eta,\phi,\psi)^{(1)} + \varepsilon^2(\eta,\phi,\psi)^{(2)} + \cdots,$$

into equations (3.12)–(3.14) leads to the $O(1)$ problems, respectively,

$$(U^2\partial_{\theta\theta}-\mu^2)\eta^{(0)} + 2[\eta^{(0)}]^3 = 0, \tag{3.17}$$

$$(U^2\partial_{\theta\theta}-\mu^2+2[\eta^{(0)}]^2)\phi^{(0)} = F_{\mathrm{I}}(\eta^{(0)}), \tag{3.18}$$

$$\psi_\theta^{(0)} = -\nu[\eta^{(0)}]^2/U, \tag{3.19}$$

and the $O(\varepsilon)$ problem associated with equation (3.12) (which is all that is needed) is

$$(U^2\partial_{\theta\theta}-\mu^2+6[\eta^{(0)}]^2)\eta^{(1)} = F_{\mathrm{R}}(\eta^{(0)},\psi^{(0)}). \tag{3.20}$$

The solutions to equations (3.17) and (3.19) are given by, respectively,

$$\eta^{(0)}(\theta, T) = \mu\,\mathrm{sech}(\mu\theta/U), \tag{3.21}$$

$$\psi^{(0)}(\theta,\tau) = -\frac{\nu\mu^2}{U}\int_{\mathrm{sgn}(U)\infty}^{\theta} \mathrm{sech}^2(\mu\xi/U)\mathrm{d}\xi = \nu\mu[1-\tanh(\mu\theta/U)]. \tag{3.22}$$

The solution for $\eta^{(0)}$ is simply the amplitude function seen in equation (3.9). The solution for $\psi^{(0)}$ is the $O(\varepsilon)$ correction to the mean flow that arises as the soliton slowly dissipates. Since the 'far field' ahead of the propagating soliton corresponds to $\mu\theta/U \to +\infty$ irrespective of the sign of U (observing that equations (2.23) and (2.24) are invariant for the transformation $A\to -A$, it is assumed, without loss of generality, that $\mu>0$), equation (3.22) ensures that $\psi^{(0)}(\theta,\tau)\to 0$ 'ahead' of the propagating soliton. However, it follows that $\psi^{(0)}(\theta,\tau)\to 2\nu\mu\neq 0$ as $\mu\theta/U\to -\infty$ (the far field *behind* the propagating soliton). Thus, a dissipation-induced 'shelf region' has emerged in the $O(\varepsilon)$ correction for the mean flow (Knickerbocker & Newell 1980). The emergence of the shelf region is a consequence of the fact that the $O(\varepsilon)$ adiabatic mean flow solution is not satisfying the integrated mass balance relation associated with equation (2.24). From the viewpoint of the asymptotics, what is required is to describe the transition back to zero in $\psi^{(0)}(\theta,\tau)$ as $\mu\theta/U\to -\infty$. This is done later in this section when equation (3.19) is reconsidered as an initial-value problem with respect to the original (X, T) variables. As will be shown, however, no shelf region emerges in the adiabatic solution for $\eta^{(0,1)}$ and the expansion is uniformly valid.

Observing that the operators on the l.h.s. of equations (3.18) and (3.20) are self-adjoint and that

$$(U^2\partial_{\theta\theta} - \mu^2 + 2[\eta^{(0)}]^2)\eta^{(0)} = 0, \tag{3.23}$$

$$(U^2\partial_{\theta\theta} - \mu^2 + 6[\eta^{(0)}]^2)\eta^{(0)}_{\theta} = 0, \tag{3.24}$$

imply that, necessarily,

$$\int_{-\infty}^{\infty} \eta^{(0)} F_{\mathrm{I}}(\eta^{(0)})\mathrm{d}\theta = 0, \tag{3.25}$$

$$\int_{-\infty}^{\infty} \eta^{(0)}_{\theta} F_{\mathrm{R}}(\eta^{(0)}, \psi^{(0)})\mathrm{d}\theta = 0. \tag{3.26}$$

The solvability conditions (3.25) and (3.26) yield transport equations that determine the slow time evolution of $\mu(\tau)$ and $U(\tau)$, as these parameters adjust to $\sigma(\tau)$ and the presence of the dissipation. The evaluation of equations (3.25) and (3.26) is straightforward (see a related calculation in Swaters 2007). Here, the transport equations will be obtained by fast phase-averaged conservation balances.

(*a*) *Determination of the transport equations from phase-averaged conservation relations*

The form of the adiabatic solutions (3.10) and (3.11) suggests the introduction of $B=\varepsilon\psi$ and $\nu=\varepsilon\hat{\nu}$ into equations (2.23) and (2.24), yielding (after dropping the caret)

$$(\partial_{TT} - \sigma)A - \mathrm{i}A_X + 2|A|^2 A = -2\varepsilon(\nu\partial_T + \psi)A + O(\varepsilon^2), \tag{3.27}$$

$$\psi_T - \nu|A|^2 = -\varepsilon\nu\psi. \tag{3.28}$$

The mass, momentum and energy conservation balances associated with equation (3.27) are given by, respectively,

$$(A^* A_T - AA_T^*)_T + \mathrm{i}(|A|^2)_X = -2\varepsilon\nu(A^* A_T - AA_T^*) + O(\varepsilon^2), \tag{3.29}$$

$$\begin{aligned}(A_X^* A_T + A_X A_T^*)_T &+ (|A|^4 + \sigma|A|^2 - |A_T|^2)_X \\ &= -2\varepsilon\psi\big(|A|^2\big)_X - 2\varepsilon\nu(A_X^* A_T + A_X A_T^*) + O(\varepsilon^2),\end{aligned} \tag{3.30}$$

$$\begin{aligned}&\left[|A|^4 + \sigma|A|^2 + |A_T|^2 + \frac{\mathrm{i}}{2}(A^* A_X - AA_X^*)\right]_T + \frac{\mathrm{i}}{2}(A^* A_T - AA_T^*)_X \\ &= \sigma_T|A|^2 - 2\varepsilon\psi(|A|^2)_T - 4\varepsilon\nu|A_T|^2 + O(\varepsilon^2),\end{aligned} \tag{3.31}$$

where $(\cdot)^*$ is the complex conjugate of $(\cdot)$. Equations (3.29), (3.30) and (3.31) are obtained by forming $A^*\times(3.27) - A\times(3.27)^*$, $A_X^*\times(3.27) + A_X\times(3.27)^*$, $A_T^*\times(3.27) + A_T\times(3.27)^*$, respectively, assuming ψ is real. The densities and fluxes are identical in form to those for the NLS equation with X and T interchanged (Grimshaw 1979*b*).

Generically, these conservation relations are of the form

$$E_T + H_X = \varepsilon G, \tag{3.32}$$

where E, H and G are the appropriate density, flux and source term, respectively. If the multiple scale *ansatz* (3.5), together with the asymptotic expansion

$$(A,\psi) \simeq (A,\psi)^{(0)} + \varepsilon(A,\psi)^{(1)} + \varepsilon^2(A,\psi)^{(2)} + \cdots,$$

is substituted into equation (3.32), expressions of the form

$$(-U\partial_\theta + \varepsilon\partial_\tau)(E^{(0)} + \cdots) + (H^{(0)} + \cdots)_\theta = \varepsilon(G^{(0)} + \cdots), \tag{3.33}$$

are obtained, which if integrated with respect to θ (assuming E, H and G all vanish sufficiently rapidly as $\theta \to \pm\infty$) yields the leading-order, phase-averaged conservation relation

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\int_{-\infty}^{\infty} E^{(0)}\mathrm{d}\theta = \int_{-\infty}^{\infty} G^{(0)}\mathrm{d}\theta. \tag{3.34}$$

It follows from equation (3.10) that

$$A \simeq \eta^{(0)}\exp(\mathrm{i}\Psi) + O(\varepsilon) \quad \text{where} \quad \Psi \equiv \left(\sigma + \mu^2 - \frac{1}{4U^2}\right)\theta + \Phi, \tag{3.35}$$

which in turn implies

$$A_T \simeq -UA_\theta^{(0)} + \mathrm{i}\Phi_T A^{(0)} + O(\varepsilon) = -UA_\theta^{(0)} + \mathrm{i}\,UA^{(0)}\left(\sigma + \mu^2 + \frac{1}{4U^2}\right) + O(\varepsilon)$$

$$= \left(-U\eta_\theta^{(0)} + \frac{\mathrm{i}\eta^{(0)}}{2U}\right)\exp(\mathrm{i}\Psi) + O(\varepsilon), \tag{3.36}$$

$$A_X \simeq A_\theta^{(0)} + O(\varepsilon) = \left[\eta_\theta^{(0)} + \mathrm{i}\left(\sigma + \mu^2 - \frac{1}{4U^2}\right)\eta^{(0)}\right]\exp(\mathrm{i}\Psi) + O(\varepsilon). \tag{3.37}$$

Thus, the leading-order mass, momentum and energy densities are given by, after a little algebra, respectively,

$$E_{\text{mass}}^{(0)} = 2\mathrm{i}\,\mathrm{Im}[(A^* A_T)^{(0)}] = \frac{\mathrm{i}[\eta^{(0)}]^2}{U}, \tag{3.38}$$

$$E_{\text{momentum}}^{(0)} = 2\,\mathrm{Re}\Big[(A_X^* A_T)^{(0)}\Big] = -2U\Big[\eta_\theta^{(0)}\Big]^2 + \left(\sigma + \mu^2 - \frac{1}{4U^2}\right)\frac{[\eta^{(0)}]^2}{U}, \tag{3.39}$$

$$E_{\text{energy}}^{(0)} = \left[|A_T|^2 + |A|^4 + \frac{\mathrm{i}}{2}(A^* A_X - AA_X^*)\right]^{(0)} = \frac{[\eta^{(0)}]^2}{2U^2}. \tag{3.40}$$

Similarly, it follows from equations (3.29)–(3.31) that

$$G_{\text{mass}}^{(0)} = -2\nu E_{\text{mass}}^{(0)}, \tag{3.41}$$

$$G_{\text{momentum}}^{(0)} = -2\psi^{(0)}([\eta^{(0)}]^2)_\theta - 2\nu E_{\text{momentum}}^{(0)}, \tag{3.42}$$

$$G_{\text{energy}}^{(0)} = \sigma_\tau[\eta^{(0)}]^2 + 2U\psi^{(0)}([\eta^{(0)}]^2)_\theta - 4\nu\left[\left(U\eta_\theta^{(0)}\right)^2 + \left(\frac{\eta^{(0)}}{2U}\right)^2\right]. \tag{3.43}$$

The integrated transport relation (3.34) associated with the mass, momentum and energy conservation relations are not independent (Swaters 2007). That is, for example, equation (3.34) evaluated for the pairs $(E_{\text{mass}}^{(0)}, G_{\text{mass}}^{(0)})$ and $(E_{\text{momentum}}^{(0)}, G_{\text{momentum}}^{(0)})$, respectively, imply that equation (3.34) evaluated for $(E_{\text{energy}}^{(0)}, G_{\text{energy}}^{(0)})$ will be satisfied. In addition, equation (3.34) evaluated for the mass and momentum density and source terms, respectively, are identical to the solvability conditions (3.25) and (3.26).

Evaluation of the phase-averaged conservation balance (3.34) for the mass and momentum density and source terms, respectively, yields

$$\mu_\tau = -2\nu\mu \Rightarrow \mu(\tau) = \mu_0\exp(-2\nu\tau), \tag{3.44}$$

$$\left(\sigma - \frac{1}{4U^2}\right)_\tau = 0 \Rightarrow U(\tau) = \frac{U_0}{\sqrt{1 + 4U_0^2[\sigma(\tau) - \sigma_0]}}, \tag{3.45}$$

where $\mu_0 = \mu(0)$; $U_0 = U(0)$; and $\sigma_0 = \sigma(0)$. The amplitude parameter decays exponentially due to the presence of the dissipation but is unaffected by the time variability in the background flow. On the other hand, the translation velocity evolves only in response to the time variability in the background flow and is unaffected by the dissipation. The fact that the dissipation does not modulate the UNS soliton velocity is identical to the adiabatic solution for the dissipating NLS soliton (Kodama & Ablowitz 1980).

The UNS soliton cannot, however, undergo adiabatic modulation for an arbitrary time variation in the background flow. It follows from the (assumed bounded) solutions (3.10) and (3.21) that $U(\tau)$ must remain real-valued. Thus, it follows from equation (3.45) that

$$\sigma(\tau) > \sigma_0 - \frac{1}{4U_0^2}. \tag{3.46}$$

The inequality in equation (3.46) cannot, in general, be inclusive since should there exist τ^* for which $\sigma(\tau^*) = \sigma_0 - 1/(4U_0^2)$, it follows from equation (3.45) that

$$\lim_{\tau \uparrow \tau^*} U(\tau) = \operatorname{sgn}(U_0)\infty,$$

which when introduced into equations (3.10) and (3.41) would suggest unphysical behaviour for $A^{(0)}$ as $\tau \to \tau^*$.

However, special cases for $\sigma(\tau)$ can be constructed in which $\lim_{\tau\to\tau^*} A^{(0)}$ does seem to make physical sense (even if no further evolution for $\tau > \tau^*$ is possible). One such special case is $\nu = 0$ (so $\mu \equiv \mu_0$), $\sigma_0 > 0$ (the marginally unstable case) and $\sigma(\tau) = \sigma_0 - \tau/(4U_0^2)$ (the background current is de-accelerating and becoming 'more stable'). In this case, $\tau^* = 1$. It follows from equation (3.45) that $U(\tau) = U_0/\sqrt{1-\tau}$ and the integrals in equations (3.5) and (3.6) can be explicitly evaluated and imply that θ and Φ exist as $\tau \to \tau^*$. Hence, the exponential term in equation (3.10) will have bounded oscillations with finite frequency as $\tau \to \tau^*$. In addition, it will follow that $\lim_{\tau\to\tau^*} \eta^{(0)} = \mu_0$. Thus, to leading order in this special case, $\lim_{\tau\to\tau^*} A^{(0)}$ exists. Generally, $\lim_{\tau\to\tau^*} A^{(0)}$ will not exist when the integrals in equations (3.5) and (3.6) do not exist as $\tau \to \tau^*$ (i.e. typically they would become unbounded) and whether or not this occurs is dependent on the rate at which $\sigma \to \sigma_0 - 1/(4U_0^2)$ as $\tau \to \tau^*$.

(*b*) *Determination of the $O(\varepsilon)$ fields*

With the transport relations (3.44) and (3.45) determined, it follows that equations (3.18) and (3.20) can be rewritten, respectively, in the form

$$(U^2\partial_{\theta\theta} - \mu^2 + 2[\eta^{(0)}]^2)\phi^{(0)} = (\nu/U - U\sigma_\tau - 4\nu U\mu^2)\left(\eta^{(0)} + 2\theta\eta_\theta^{(0)}\right), \tag{3.47}$$

$$(U^2\partial_{\theta\theta} - \mu^2 + 6[\eta^{(0)}]^2)\eta^{(1)} = -2(\psi^{(0)} + 2\nu\mu^2\theta/U)\eta^{(0)} + 2U(U^2\sigma_\tau - 3\nu)\eta_\theta^{(0)} + 4U\theta(U^2\sigma_\tau - \nu)\eta_{\theta\theta}^{(0)}. \tag{3.48}$$

It is straightforward to verify that the r.h.s. of equations (3.47) and (3.48) satisfy, respectively,

$$\int_{-\infty}^{\infty} \eta^{(0)} \times (\nu/U - U\sigma_\tau - 4\nu U\mu^2)\left(\eta^{(0)} + 2\theta\eta_\theta^{(0)}\right)\mathrm{d}\theta$$

$$= \int_{-\infty}^{\infty} \eta_\theta^{(0)} \times \Big(-2(\psi^{(0)} + 2\nu\mu^2\theta/U)\eta^{(0)} + 2U(U^2\sigma_\tau - 3\nu)\eta_\theta^{(0)} + 4U\theta(U^2\sigma_\tau - \nu)\eta_{\theta\theta}^{(0)}\Big)\mathrm{d}\theta = 0,$$

i.e. the solvability conditions (3.25) and (3.26) hold for all differentiable $\sigma(\tau)$.

Introducing the transformations

$$\phi^{(0)} = \eta^{(0)}\tilde{\phi} \quad \text{and} \quad \eta^{(1)} = \eta_\theta^{(0)}\tilde{\eta},$$

into equations (3.47) and (3.48) leads to, respectively,

$$U^2\partial_\theta([\eta^{(0)}]^2\tilde{\phi}_\theta) = (\nu/U - U\sigma_\tau - 4\nu U\mu^2)(\theta[\eta^{(0)}]^2)_\theta, \tag{3.49}$$

$$U^2\partial_\theta\left(\left[\eta_\theta^{(0)}\right]^2\tilde{\eta}_\theta\right) = \nu\mu^2[\eta^{(0)}]^2/U - 3\nu U\left[\eta_\theta^{(0)}\right]^2 + \left\{2U(U^2\sigma_\tau - \nu)\left[\eta_\theta^{(0)}\right]^2 - (\psi^{(0)} + 2\nu\mu^2\theta/U)\left[\eta^{(0)}\right]^2 + \right\}_\theta, \tag{3.50}$$

which can be integrated twice to yield

$$\phi^{(0)} = \frac{\mu}{2}(\nu/U - U\sigma_\tau - 4\nu U\mu^2)(\theta/U)^2\mathrm{sech}(\mu\theta/U), \tag{3.51}$$

$$\eta^{(1)} = \{\nu[1 + 2\mu\theta/U][(\mu\theta/U)\tanh(\mu\theta/U) - 1] - \sigma_\tau\mu^2\theta^2\tanh(\mu\theta/U)\}\tanh(\mu\theta/U)\mathrm{sech}^2(\mu\theta/U). \tag{3.52}$$

It follows from equations (3.51) and (3.52) that $\phi^{(0)}$ and $\eta^{(1)}$ are defined for all $\theta\in\mathbb{R}$ and that $|\phi^{(0)}, \eta^{(0)}|\to 0$ exponentially and rapidly as $|\theta|\to\infty$. Thus, no shelf region develops in these $O(\varepsilon)$ solutions in the far field behind the soliton.

The shelf region in $\psi^{(0)}$ that arises in the adiabatic solution (3.22) is removed by solving the initial-value problem associated with equation (3.19). With respect to the original (X, T) variables, equation (3.19) is given by

$$\psi_T^{(0)}(X, T) = \nu[\eta^{(0)}]^2. \tag{3.53}$$

The spatial non-uniformity in equation (3.22) as $X\to-\infty$ develops because the adiabatic solution does not satisfy the integrated mass balance relation associated with equation (3.53) given by

$$\partial_T\int_{-\infty}^{\infty}\psi^{(0)}(X, T)\mathrm{d}X = \nu\int_{-\infty}^{\infty}[\eta^{(0)}]^2\mathrm{d}X = 2\nu\mu U, \tag{3.54}$$

where $\eta^{(0)}$ has been written in the form

$$\eta^{(0)} = \mu \operatorname{sech}\left[\frac{\mu}{U}\left(X - \frac{1}{\varepsilon}\int_0^{\varepsilon T} U(\xi)\mathrm{d}\xi\right)\right].$$

If equation (3.22) is substituted into the integral in the l.h.s. of equation (3.54), it follows that

$$\int_{-\infty}^{\infty} \psi^{(0)}(X, T)\mathrm{d}X = -\infty,$$

(assuming $\nu\mu > 0$) which clearly shows that the balance equation (3.54) is not satisfied by the adiabatic solution for $\psi^{(0)}$.

The spatial non-uniformity in $\psi^{(0)}$ as $X \to -\infty$ is removed and the integrated mass balance equation (3.54) will be satisfied by solving equation (3.53) as an initial-value problem subject to $\psi^{(0)}(X, 0) = 0$. Denoting this solution as $\psi^{(0)}_{\mathrm{IVP}}(X, T)$, it follows that

$$\psi^{(0)}_{\mathrm{IVP}}(X, T) = \nu\mu\left\{\tanh(\mu X/U) - \tanh\left[\frac{\mu}{U}\left(X - \frac{1}{\varepsilon}\int_0^{\varepsilon T} U(\xi)\mathrm{d}\xi\right)\right]\right\}, \tag{3.55}$$

where terms of $O(\varepsilon)$ have been ignored. It is clear that equation (3.55) satisfies

$$\lim_{|X|\to\infty} \psi^{(0)}_{\mathrm{IVP}}(X, T) = 0,$$

$\forall\ T \geq 0$ so that there is no shelf region and, consequently, the resulting solution is not exponentially non-uniform with respect to X. Moreover, after a little algebra, it can be shown that

$$\int_{-\infty}^{\infty} \psi^{(0)}_{\mathrm{IVP}}(X, T)\mathrm{d}X = \nu U \ln\left|\frac{\cosh\left(\frac{1}{\varepsilon}\int_0^{\varepsilon T} U(\xi)\mathrm{d}\xi\right) + \sinh\left(\frac{1}{\varepsilon}\int_0^{\varepsilon T} U(\xi)\mathrm{d}\xi\right)}{\cosh\left(\frac{1}{\varepsilon}\int_0^{\varepsilon T} U(\xi)\mathrm{d}\xi\right) - \sinh\left(\frac{1}{\varepsilon}\int_0^{\varepsilon T} U(\xi)\mathrm{d}\xi\right)}\right|,$$

from which it follows that

$$\partial_T \int_{-\infty}^{\infty} \psi^{(0)}_{\mathrm{IVP}}(X, T)\mathrm{d}X = 2\nu\mu U + O(\varepsilon),$$

so that the integrated mass balance equation (3.54) will be satisfied.

Finally, the adiabatic solution (3.22) is recovered by writing $\psi^{(0)}_{\mathrm{IVP}}$ in terms of the adiabatic variables (θ, τ), which is given by

$$\psi^{(0)}_{\mathrm{IVP}} = \nu\mu\left\{\tanh\left[\frac{\mu}{U}\left(\theta + \frac{1}{\varepsilon}\int_0^{\tau} U(\xi)\mathrm{d}\xi\right)\right] - \tanh(\mu\theta/U)\right\},$$

and observing that in the limit where θ and τ are $O(1)$ and $\varepsilon \to 0$, it follows that

$$\tanh\left[\frac{\mu}{U}\left(\theta + \frac{1}{\varepsilon}\int_o^{\tau} U(\xi)\mathrm{d}\xi\right)\right] \simeq 1.$$

4. Conclusions

Baroclinic FG models correspond to a dynamical regime describing the low-frequency evolution of stratified ocean currents with large amplitude isopycnal deflections on length scales longer than the internal deformation radius.

In particular, FG theory can describe surface ocean currents in which the isopycnals intersect the surface as they do in frontal regions and warm-core rings and eddies.

The principal purpose of this paper has been to derive and analyse a finite-amplitude model describing the evolution of marginal solitary disturbances for a baroclinic FG model when a time-varying background flow and dissipation are present. Understanding the effects of time variability in the transition problem for ocean currents has the potential for dramatically altering the classical viewpoint of the relationship and interaction between the eddy and mean flow fields. For example, the periodic perturbing of a coastal current by the tides may lead to destabilization even if the coastal current is instantaneously stable at each moment in time. On the other hand, and in the other direction, long time-scale periodic, e.g. seasonal, forcing may act to stabilize an otherwise unstable large-scale current. The implications of these types of processes on the planetary-scale mixing of the oceans may make an important contribution to natural climate variability.

It has been shown that with dissipation and time variability present in the transition problem for a marginal FG flow, the amplitude of the fundamental normal mode and its accompanying mean flow satisfy a two-equation extension of the UNS equation. The UNS equation can be thought of as the classical NLS equation but with the space and time variables interchanged. More properly, the UNS equation corresponds to a special limit of the Ginzburg–Landau equation.

In the inviscid limit with no background time variability, the wave-packet model possesses a steadily travelling oscillating solitary wave solution. This solution corresponds to a robust surface-intensified isolated anomaly that is capable of transporting water properties over large distances. A multiple scale asymptotic analysis was developed to describe the effect of time variability and dissipation on the solitary wave solution. The asymptotic theory is based on a direct perturbation approach and fast phase-averaged conservation relations. In addition to determining the adiabatic evolution of the deforming soliton, the first-order perturbation fields are also obtained. Specifically, it was shown that the soliton amplitude parameter decays exponentially due to the presence of the dissipation but is unaffected by the time variability in the background flow. On the other hand, the soliton translation velocity is unaffected by the dissipation and evolves only in response to the time variability in the background flow.

There is no shelf region that emerges in the asymptotic solution for the normal mode amplitude (i.e. the perturbation solution is uniformly valid spatially). A dissipation-generated shelf region does emerge in the asymptotic adiabatic solution for the induced mean flow. It was shown that by considering the proper initial-value problem for the first-order perturbation mean flow field, the spatial non-uniformity is eliminated. The adiabatic solution was recovered as the appropriate asymptotic limit of the initial-value solution.

Preparation of this paper was supported in part by the Natural Sciences and Engineering Research Council of Canada. Email: gordon.swaters@ualberta.ca, URL: http://pacific.math.ualberta.ca/gordon/

References

Craik, A. D. D. 1985 *Wave interactions and fluid flows*. Cambridge, UK: Cambridge University Press.

Cushman-Roisin, B. 1986 Frontal geostrophic dynamics. *J. Phys. Oceanogr.* **22**, 117–127. (doi:10.1175/1520-0485(1992)022<0117:TLGDPI>2.0.CO;2)

Cushman-Roisin, B., Sutyrin, G. G. & Tang, B. 1992 Two-layer geostrophic dynamics. Part I: governing equations. *J. Phys. Oceanogr.* **22**, 758–768. (doi:10.1175/1520-0485(1992)022<0117:TLGDPI>2.0.CO;2)

Gibbon, J. D. & McGuinness, M. J. 1981 Amplitude equations at the critical points of unstable dispersive physical systems. *Proc. R. Soc. A* **377**, 185–219. (doi:10.1098/rspa.1981.0121)

Gibbon, J. D., James, I. N. & Moroz, I. M. 1979 An example of soliton behavior in a rotating baroclinic fluid. *Proc. R. Soc. A* **367**, 219–237. (doi:10.1098/rspa.1979.0084)

Grimshaw, R. 1979*a* Slowly varying solitary waves. I. Korteweg–de Vries equation. *Proc. R. Soc. A* **368**, 359–375. (doi:10.1098/rspa.1979.0135)

Grimshaw, R. 1979*b* Slowly varying solitary waves, II. Nonlinear Schrödinger equation. *Proc. R. Soc. A* **368**, 377–388. (doi:10.1098/rspa.1979.0136)

Ha, S.-J. & Swaters, G. E. 2006 Finite-amplitude baroclinic instability of time-varying abyssal currents. *J. Phys. Oceanogr.* **36**, 122–139. (doi:10.1175/JPO2838.1)

Huang, N.-N., Chi, S., Lou, B. L. & Chen, X. J. 2000 A direct theory for the perturbed unstable nonlinear Schrödinger equation. *J. Math. Phys.* **41**, 2931–2942. (doi:10.1063/1.533281)

Iizuka, T. & Wadati, M. 1990 The Rayleigh–Taylor instability and nonlinear waves. *J. Phys. Soc. Jpn.* **59**, 3182–3192. (doi:10.1143/JPSJ.59.3182)

Karsten, R. H. & Swaters, G. E. 1996 A note on the stability theory of buoyancy-driven ocean currents over a sloping bottom. *Z. Angew. Math. Phys.* **47**, 28–38. (doi:10.1007/BF00917572)

Karsten, R. H. & Swaters, G. E. 2000 Nonlinear effects in two-layer large-amplitude geostrophic dynamics. Part 2. The weak-beta case. *J. Fluid. Mech.* **412**, 161–196. (doi:10.1017/S0022112000008272)

Kaup, D. J. & Newell, A. C. 1978 Solitons as particles, oscillators and in slowly changing media: a singular perturbation theory. *Proc. R. Soc. A* **361**, 413–446. (doi:10.1098/rspa.1978.0110)

Klein, P. & Pedlosky, J. 1992 The role of dissipation mechanisms in the nonlinear dynamics of unstable baroclinic waves. *J. Atmos. Sci.* **49**, 29–48. (doi:10.1175/1520-0469(1992)049<0029:TRODMI>2.0.CO;2)

Knickerbocker, C. J. & Newell, A. C. 1980 Shelves and the Korteweg–de Vries equation. *J. Fluid Mech.* **98**, 803–818. (doi:10.1017/S0022112080000407)

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

McWilliams, J. C. 1984 The emergence of isolated coherent vortices in turbulent flow. *J. Fluid Mech.* **146**, 21–43. (doi:10.1017/S0022112084001750)

Mooney, C. J. & Swaters, G. E. 1996 Finite amplitude baroclinic instability of a mesoscale gravity current in a channel. *Geophys. Astrophys. Fluid Dyn.* **82**, 173–205. (doi:10.1080/03091929608213634)

Pedlosky, J. 1972 Finite amplitude baroclinic wave packets. *J. Atmos. Sci.* **29**, 680–686. (doi:10.1175/1520-0469(1972)029<0680:FABWP>2.0.CO;2)

Pedlosky, J. 1987 *Geophysical fluid dynamics*. New York, NY: Springer.

Pedlosky, J. & Thomson, J. 2003 Baroclinic instability of time dependent currents. *J. Fluid Mech.* **490**, 189–215. (doi:10.1017/S0022112003005007)

Poulin, F. J., Flierl, G. R. & Pedlosky, J. 2003 Parametric instability in oscillatory shear flow. *J. Fluid Mech.* **481**, 329–353. (doi:10.1017/S0022112003004051)

Reszka, M. K. 1997 Finite amplitude waves and eddy development on a baroclinically unstable front over a sloping bottom. M.Sc. thesis, University of Alberta, Edmonton.

Reszka, M. K. & Swaters, G. E. 1999*a* Eddy formation and interaction in a baroclinic frontal geostrophic model. *J. Phys. Oceanogr.* **29**, 3025–3042. (doi:10.1175/1520-0485(1999)029<3025:EFAIIA>2.0.CO;2)

Reszka, M. K. & Swaters, G. E. 1999*b* Numerical investigation of baroclinic instability in the Gaspé current using a frontal geostrophic model. *J. Geophys. Res.* **104**, 25 685–25 696. (doi:10.1029/1999JC900189)

Reszka, M. K. & Swaters, G. E. 2004 Evolution of initially axisymmetric buoyancy jets: a numerical study. *J. Fluid Mech.* **501**, 355–377. (doi:10.1017/S0022112003007535)

Swaters, G. E. 1993 On the baroclinic dynamics, Hamiltonian formulation and general stability characteristics of density-driven surface currents and fronts over a sloping continental shelf. *Phil. Trans. R. Soc. A* **345**, 295–325. (doi:10.1098/rsta.1993.0131)

Swaters, G. E. 2000 *Introduction to Hamiltonian fluid dynamics and stability theory*. Monographs and surveys in pure and applied mathematics. Boca Raton, FL: Chapman & Hall/CRC Press.

Swaters, G. E. 2006 The meridional flow of source-driven abyssal currents in a stratified basin with topography. Part I. Model development and dynamical properties. *J. Phys. Oceanogr.* **36**, 335–355. (doi:10.1175/JPO2855.1)

Swaters, G. E. 2007 Perturbations of soliton solutions to the unstable nonlinear Schrödinger and sine-Gordon equations. *Stud. Appl. Math.* **118**, 99–116. (doi:10.1111/j.1467-9590.2007.00365.x)

Tan, B. & Liu, S. 1995 Collision interactions of solitons in a baroclinic atmosphere. *J. Atmos. Sci.* **52**, 1501–1512. (doi:10.1175/1520-0469(1995)052<1501:CIOSIA>2.0.CO;2)

Tang, B. & Cushman-Roisin, B. 1992 Two-layer geostrophic dynamics. Part II: geostrophic turbulence. *J. Phys. Oceanogr.* **22**, 128–138. (doi:10.1175/1520-0485(1992)022<0128:TLGDPI>2.0.CO;2)

Turnbull, M. R. 2006 Finite-amplitude instability of time-varying frontal geostrophic currents. M.Sc. thesis, University of Alberta, Edmonton.

Yajima, N. & Tanaka, M. 1988 Soliton modes in an unstable plasma. *Suppl. Prog. Theor. Phys.* **94**, 138–162.

Yajima, N. & Wadati, M. 1990*a* Soliton solution and its property of unstable nonlinear Schrödinger equation. *J. Phys. Soc. Jpn.* **59**, 41–47. (doi:10.1143/JPSJ.59.41)

Yajima, N. & Wadati, M. 1990*b* Solitons in electron beam plasma. *J. Phys. Soc. Jpn.* **59**, 3237–3248. (doi:10.1143/JPSJ.59.3237)